18th International Congress of Phonetic Sciences

Satellite Event

The Evolution of Phonetic Capabilities:

Causes
Constraints
Consequences

Glasgow, August 12, 2015

Proceedings edited by
Hannah Little

TABLE OF CONTENTS

INTRODUCTION TO THE EVOLUTION OF PHONETIC CAPABILITIES SATELLITE MEETING

Hannah Little

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

1. AIMS OF THE MEETING

In the field of language evolution, new techniques have emerged in computational and mathematical modelling, experimental paradigms, brain and vocal tract imaging, and typological data. These techniques have allowed us to address questions relevant to the evolution of our phonetic capabilities. Multidisciplinary discourse is now necessary.

Previously, the evolution of speech has been addressed, for the most part, separately to the evolution of language. The difference between these two fields can be characterised in the difference between language and speech, language being a communication system irrespective of its modality, while speech refers only to the spoken modality. Fitch [2] argues that while the evolution of speech and language are obviously very relevant to each other, they might have been subject to different evolutionary pressures, making it advantageous to investigate them separately. Indeed, within language evolution, language is framed as the result of cognitive mechanisms, whether specific to language or not, or functional mechanisms such as communication or transmission. Studies on the evolution of speech, on the other hand, focus primarily on the evolutionary pressures behind the physical biology of the vocal tract (see [1] for a review). This has lead to work on the evolution of speech focussing on biological evolution, while work on the evolution of language focusses more on linguistic development and cultural evolution. In the interests of opening a dialogue between language evolution and speech evolution, we will consider how the physical aspects of a linguistic modality might shape our language, and how our phonetic capabilities at the speech level may influence our phonology at the language level.

2. CONTRIBUTIONS

This workshop has a wide range of contributions, from across disciplines, which cover typological data and acquisition data, as well as simulations and models. All of the contributions explore evolutionary pressures causing the emergence of our phonetic or phonological capabilities, both in biological and cultural evolution. Further, they discuss the consequences that biological constraints, or even external constraints in our environments, might have on processes of cultural evolution, and vice versa.

Firstly we have contributions which deal directly with physical aspects of the vocal tract affecting cultural processes. John H. Esling, Allison Benner and Scott R. Moisik investigate how the larynx affects language acquisition, and what this might mean for both the biological and cultural evolution of language. Scott R. Moisik and Dan Dediu created a vocal tract model to test if anatomical variation can bias the production of clicks, in order to explore the consequences that physical characteristics might have on cultural evolution. Padraic Monaghan and Willem H. Zuidema also touch on how constraints of speech production might affect language, presenting a corpus study investigating effects on repetition of phonemes within words.

Seán G. Roberts, Caleb Everett and Damián Blasi introduce “evolutionary geophonetics", an exploration on whether climate might influence language evolution through either biological or cultural adaptations.

Finally, we have contributions which use computational models to address the topics of the meeting. Bodo Winter and Andy Wedel present a simulation showing that functional pressions for distinctiveness at the word level, can influence variation at the phonetic level. Bill Thompson uses his model to investigate how variation in phonetic categories at the population-level can be affected by individual-level category acquisition.

3. REFERENCES

[1] Fitch, W. T. 2000. The evolution of speech: a comparative review. *Trends in cognitive sciences* 4(7), 258–267.

[2] Fitch, W. T. 2002. Comparative vocal production and the evolution of speech: reinterpreting the descent of the larynx. Wray, A., (ed), *The transition to language*. Oxford University Press Oxford 21–45.

LARYNGEAL ARTICULATORY FUNCTION AND SPEECH ORIGINS

John H. Esling[1], Allison Benner[1], Scott R. Moisik[2]

[1]University of Victoria, [2]Max Planck Institute for Psycholinguistics, Nijmegen
esling@uvic.ca, abenner@uvic.ca, Scott.Moisik@mpi.nl

ABSTRACT

The larynx is the essential articulatory mechanism that primes the vocal tract. Far from being only a glottal source of voicing, the complex laryngeal mechanism entrains the ontogenetic acquisition of speech and, through coarticulatory coupling, guides the production of oral sounds in the infant vocal tract. As such, it is not possible to speculate as to the origins of the speaking modality in humans without considering the fundamental role played by the laryngeal articulatory mechanism. The Laryngeal Articulator Model, which divides the vocal tract into a laryngeal component and an oral component, serves as a basis for describing early infant speech and for positing how speech sounds evolving in various hominids may be related phonetically. To this end, we offer some suggestions for how the evolution and development of vocal tract anatomy fit with our infant speech acquisition data and discuss the implications this has for explaining phonetic learning and for interpreting the biological evolution of the human vocal tract in relation to speech and speech acquisition.

Keywords: laryngeal, larynx, vocal tract anatomy, infant speech, ontogeny

1. INTRODUCTION

The ‘laryngeal articulator,’ consisting of the glottal mechanism, the supraglottic epilaryngeal tube, the pharyngeal/epiglottal mechanism, and including three levels of folds – the vocal folds, the ventricular folds, and the aryepiglottic folds – is responsible for the generation of multiple source vibrations and for the complex modification of the epilaryngeal and pharyngeal resonating chambers that account for a wide range of contrastive auditory qualities. These qualities are observed in a surprisingly large number of the languages of the world, both linguistically and paralinguistically, and they account for sounds labelled in the IPA as ‘pharyngeal’ and ‘epiglottal,’ as various phonation types, as tonal register phonatory contrasts, or as vowel harmony secondary qualities. They reflect an expanding range of what have been known as the ‘states of the glottis’ (now more properly termed ‘states of the larynx’) [9, 14, 8, 23]. The laryngeal mechanism constitutes a significantly large and strategic portion of the vocal tract, as depicted in the ‘Laryngeal Articulator Model’ [10, 11], which has nevertheless been generally overlooked in considering the ontogeny and phylogeny of the phonetic capacity.

It has also been observed that infants, in their first months of life, produce a range of utterances, reflecting both phonatory possibilities and stricture types, that can be directly attributed to the laryngeal articulator mechanism. Systematic observation of infants’ early speech production reveals that the control of articulatory detail in the pharynx is mastered during the first year of life [3, 13, 2, 18]. The control and growing understanding of manner of articulation in the pharynx (within the laryngeal mechanism) appears to be a prerequisite for expanding articulatory control into the oral vocal tract. Taking the larynx/pharynx as a starting point for the ontogenetic learning of the speech production capacity is likely to offer productive insights into the phylogenetic development of speech.

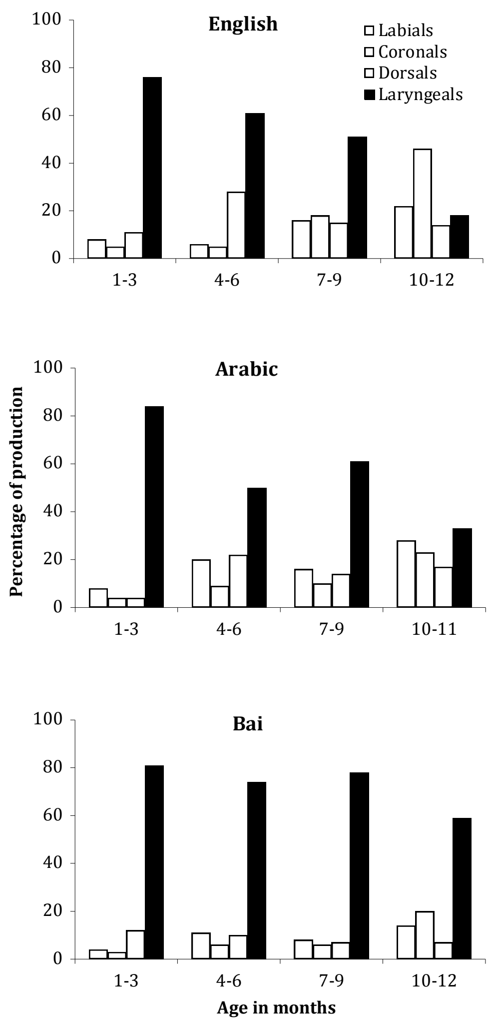
2. INFANT SPEECH ACQUISITION

2.1. Speech begins in the pharynx (with the laryngeal articulator)

Research into the earliest vocalizations by infants in English, French, Arabic, and Bai (Tibeto-Burman) contexts shows that: (1) speech begins in the pharynx, (2) the production of vocalic phonation and of consonantal stricture begins with laryngeally constricted settings, (3) infants actively explore their laryngeal phonetic production capacity through ‘dynamic alternations’ of paired contrasts, as those contrasts are discovered, and (4) infants often generate oral (lingual, labial, etc.) sounds with a primary laryngeal vocalization which precedes the oral articulation or is maintained as a coarticulation with the oral sound. Evidence from the Infant Speech Acquisition (InSpA) Project [12] illustrates instances of systematic ‘phonetic play’ that demonstrate how infants acquire basic control over the speech mechanism and the arrays of place and manner of articulation within the larynx during their first year of life.

Anatomically, laryngeal constriction is the first phonetic mechanism available to the infant, since a short (raised) and relatively flat laryngeal vocal tract is predisposed [1]. After (vocalic) crying with constricted (retracted) vowel quality, the 'first sound' that infants can be said to produce as an articulatory (consonantal) stricture is epiglottal stop [ʡ] [13, 12], which they do beginning from the first weeks of life. This stricture is a function of the laryngeal constrictor as the primary airway-protection reflex [16]. Glottal stop [ʔ], requiring more careful control than epiglottal stop, emerges later, early in the second month. Pharyngeal fricatives, approximants and trills appear early. Figure 1 shows the results of an analysis of 4,499 consonantal sounds produced by infants (English: 1,195; Arabic: 1,696; Bai: 1,608). The results clearly illustrate the prevalence of laryngeal sounds (including pharyngeal and glottal sounds) early in infancy and the increase in oral sounds throughout the first year in the production of infants from these three language groups.

Figure 1: Percentage of infants' production in terms of place of articulation according to infants' linguistic background and age group.



Chi-squared and Cramer's V analyses were performed on the consonantal data, split according to the different age groups (1-3, 4-6, 7-9, and 10-12 months) to test the strength of association between language and place of articulation for each of the four age groups. The results indicate that despite the significant association between language and place of articulation for all age groups (for all chi-squared results $p < .01$), the strength of the relationship between these two variables is very weak at 1-3 months (Cramer's V = .104), but considerably stronger at 10-12 months (Cramer's V = .239). These results suggest that as infants approach the end of their first year, their production becomes distinctive from one language group to another, presumably due to the influence of their ambient language. Early in infancy, the prevalence of laryngeal sounds illustrates our hypothesis that speech begins in the pharynx.

Similarly, phonatory configurations where laryngeal constriction dominates (harsh, whispery, and creaky voice) appear before unconstricted (modal, breathy, or falsetto) phonation. In the earliest months, laryngeally constricted production dominates in all languages observed. Analyses of an initial 3,197 utterances (English: 932; Arabic: 1,011; Bai: 1,254), contrasting only auditorily-evaluated constricted vs. unconstricted utterances across age groups, are significant (X^2 (3) = 93.34, $p < .001$), indicating that the incidence of laryngeal constriction in infants' vocalizations varies primarily as an inverse function of age, irrespective of linguistic background [1]. In all language groups, early vocalizations are overwhelmingly constricted, i.e. harsh, creaky, pharyngealized, raised-larynx, etc. As illustrated in Figure 2, the incidence of laryngeal constriction decreased progressively throughout the first year for infants from all three language groups examined, while still forming a major part of their vocal repertoire at the 10-12 month period. In summary, open-airway phonetic realizations occur only rarely until halfway through the first year. It could be said that laryngeally constricted qualities and strictures are reflexively innate, while open (less protective) qualities and strictures are learned.

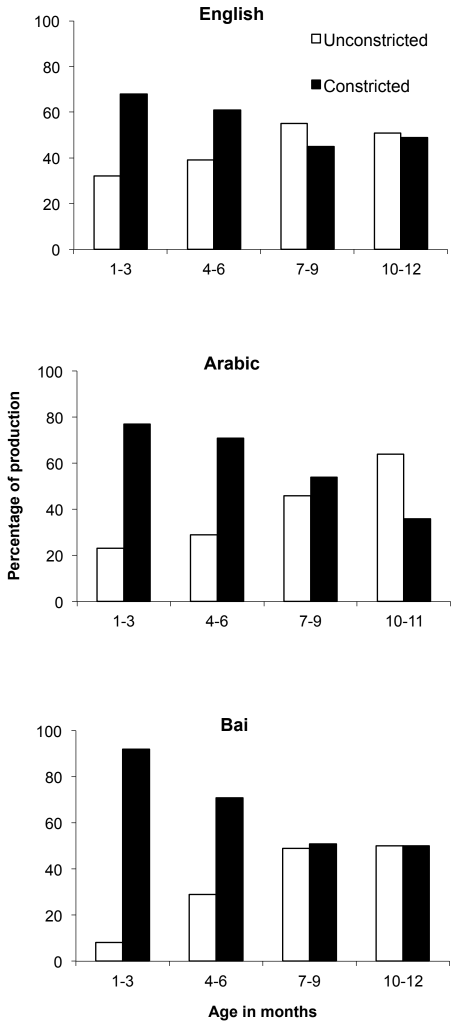


Figure 2: Constricted and unconstricted voice quality settings produced by English, Arabic, and Bai infants.

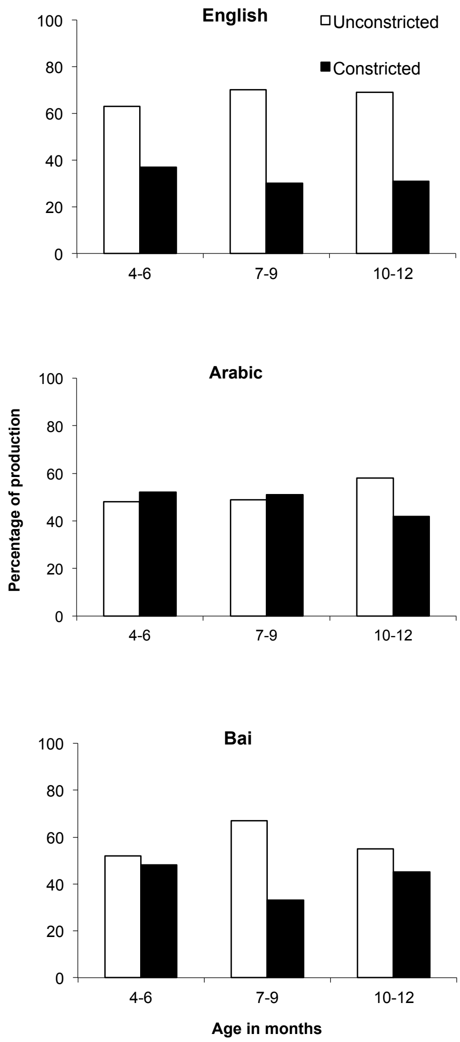


Figure 3: Constricted and unconstricted phonatory settings in the babbling of English, Arabic, and Bai infants.

2.2. Laryngeally constricted vocalization persists

Even during babbling, towards the end of the first year, when oral sounds become preferred, some constricted qualities persist, especially in those languages that contain pharyngeals (Arabic) or constricted registers (Bai) in their phonologies (Figure 3). For example, at the end of the first year, in months 10-12, only 31% of the babbling of English infants includes laryngeal constriction, compared to 42% and 45% for the Arabic and Bai infants, respectively.

Furthermore, as control over the articulators grows and oral strictures begin to be used, sounds that are learned at new oral places of articulation often occur with secondary ‘accompaniments’ from the original laryngeal articulator: coarticulatory events termed ‘pharyngeal priming.’ The preference in babbling for oral sounds may relate to the split between brain stem neural control and cortical neural control, where brain stem control can be posited to account for the reflexive emergence of the innate use of the laryngeal articulator and cortical control hypothesized to coincide with the shift from phonetic pre-babbling practice to the primarily oral control exhibited in the babbling stage.

3. EVOLUTIONARY ENTAILMENTS

In Burling’s [5] account of the evolution of language and speech, the assertion that human speech sounds have conventional meaning rather than just being iconic from an early stage is supported by our evidence from phonetic ontogeny. What our research adds to the equation is that infants acquire motor control over contrastively useful parcels of speech at a surprisingly early age and in a reflexively rich but visually hidden part of the vocal tract. Any speculation about oscillatory patterns of articulators [22] needs to take into account that these patterns would have developed in the pharynx first, before they progressed to the mouth or the jaw. This provokes speculation about early hominids. If speech sounds develop ontogenetically beginning in the pharynx, as our research has shown, then this invites the possibility that speech sounds could have developed phylogenetically in the pharynx. At the very least, the laryngeal articulator capabilities of early hominid anatomy need to be considered. In reviewing accounts of language evolution such as Burling's, it is important to recognize that the agents of acquisition and change are infants in both cases rather than adults. That is, the speech capacity did not start with an early hominid who had already reached adulthood. Speech representations in every epoch begin with infants, from day one, acquiring phonetic production capabilities in a systematic progression from the larynx/pharynx outwards. At some point in time, infants gained the awareness that their own auto-generated sounds could be used for symbolic meaning. These stimuli would for a time be reflexive, eventually if only occasionally being responded to by an adult (most likely in indirect ways) and reinforced in various directions. In our methodology, it has become clear that adults become intensely aware of the human sound-producing capability when they have infants who are generating the basic elements of phonetic motor production during the first several months of life. The elements become familiar to the adults, but the infant is the driving force; i.e. the sounds are created by each infant, in a logical progression of how sounds can be produced in the pharynx, rather than being 'taught' to the infant. That is, we all learn phonetics 'experimentally' [cf. 6].

The crux of the issue is: if contemporary infants start phonetic acquisition with the laryngeal constrictor mechanism as first articulator, then how far back along the evolutionary path has this been the case? Early hominid infants, once they had the required cognitive criteria for language development that Burling enumerates, could be expected to have generated sounds similar to those pharyngeal sounds that every infant generates today, which have the potential to represent linguistic meaning, and which the infant 'discovers' as having that potential. The mechanism for drawing the phonetic and the semantic processes together would have likely been precisely because of the infant-adult interaction. Burling's observation that 'it is the parent, not the child, who is the imitator' [5:110] is given support by our observations of each infant's autogeneration of laryngeal contrasts (at the purely phonetic level) and remarkable control and early mastery of the innate laryngeal sound-producing instrument.

4. PHYLOGENY, DISCUSSION

4.1. Anatomy and laryngeal articulation

A great deal of attention has been placed on the size of the laryngeal vocal tract in speculation about the phylogenetic substrate necessary for the emergence of speech [15, 20, 21]. The main thrust of this discussion has focused on the proportioning of the oral and pharyngeal cavities (the horizontal/vertical supralaryngeal vocal tract ratio or SVTh/SVTv) in relation to potential phonetic vowel categories and their degree of quantality, *sensu* Stevens [27]. Recently, Boë et al. have asserted the importance of forming oral consonantal stricture [4], and the suggestion of biomechanical limitations on the chimpanzee tongue has been made in favour of this account [28].

The evidence that the larynx is the first domain of phonetic exploration adds yet another degree of complexity to the question of how speech may have evolved. While the human larynx is indeed situated low within the vocal tract, the descent during ontogeny of the laryngeal cartilages relative to the hyoid bone follows a remarkably similar pattern to that observed for chimpanzees [25]. This is thus a phylogenetically old component of the anthropoid vocal tract's developmental sequence; in most other mammals, the hyo-thyroid complex remains bound together and inhibits independent lingual-laryngeal control [19]. The relatively high early position of the larynx relative to the hyoid might have a protective function during infant cry vocalizations (mostly associated with mother-infant separation in non-human primates [24]). This might operate through the action of non-linear source-filter coupling [29], which, when there is substantial epilaryngeal narrowing, serves to increase the acoustic efficiency of the vocal folds. This has the benefit of reducing vocal fold stresses during crying vocalization while still generating a vocalization sufficiently intense to attract the attention of the caregiver (cf. [23]). The naturally constricted larynx also offers other enhancements to the attention-getting function of cry through, for example, the perturbation to phonation (harshness at the vocal fold level) or the accompaniment by extraglottal vibrations associated with the epilarynx, such as those of the ventricular folds, the aryepiglottic folds, or the epiglottis.

As has been shown in our research, this predisposing positioning of the larynx relative to the hyoid bone provides the grounds for the acquisition of the first consonantal stricture (the epiglottal stop) and for the development of manner of articulation (through manipulation of stop, approximant, fricative, and trilling phonetic postures). The pre-constricted posture also has other benefits for phonetic learning. A major challenge in understanding the acquisition of the complex motor control of speech [17] is how the innumerable degrees of freedom of the articulators are mastered. Early hyo-laryngeal approximation and its constraining of infant vocalizations initially to laryngeally constricted sounds serves to reduce considerably the search space of learning the motor control mechanisms behind producing different forms of consonantal strictures. We suspect that these laryngeally enacted processes constitute an early cortical mapping for manner categories upon which oral manners can be developed.

4.2. Unlocking the oral articulators

The other essential component of phonetic behaviour is the development of oral-laryngeal coarticulation, which is critical in the formation of voicing contrasts on obstruents and is essential in the production of tonal and intonational patterns. As the human vocal tract develops [26], the horizontal (i.e. oral) component exhibits a sudden spurt of growth which then nearly halts towards the end of the second year, having attained approximately its pre-adolescence scale. By comparison, there is ongoing growth of the laryngeal vocal tract throughout early childhood, which ultimately gives rise to the characteristic separation between hyoid bone and palate. By comparison, the oral vocal tract of the chimpanzee shows a much faster rate of growth than the laryngeal vocal tract. We might suspect that these continuously changing proportions of the vocal tract would offer some difficulty to the early establishment of place of articulation categories. Whatever ultimately drove the development of a flattened facial profile in humans, we suspect it offers a great advantage for phonetic learning, at least over the chimpanzee vocal tract, by being relatively stable during the post-babbling period (during the second year of life).

It is roughly at the end of the first year, once our larynx has gone through the first crucial 7-8 months of descent in relation to the hyoid bone, that the post-laryngeal phase of phonetic learning begins. By this point we can think of the oral articulators as being ‘unlocked’. The infant now has the challenge of learning to control many more degrees of freedom for phonetic purposes but can draw on control schemes in place for functions such as suckling (control of the lips and the tongue) and swallowing (control of the lips, tongue, soft palate, and larynx) juxtaposed against the cortical setting established for the control of basic phonetic categories of manner of articulation. The vocal behaviour of our primate cousins does not seem to include or at least favour these consonantal properties, being instead characterized primarily by modulation of vowel and phonatory qualities.

5. SUMMARY

The efficacy of vocalization as a social tool is ancient in the primate clade. Humans have taken the remarkable step of exploiting vocalization for the purposes of communication, and, as the predominant modality of human language, it is hard to believe that the need to acquire and use speech did not have some selective effect on our biology. With that stated, it is also the case that those components of ontogeny relating to the position and posturing of the larynx, which we have argued are an essential component of our phonetic learning and capacity, were already in place before language appeared. It strikes us as highly plausible that hominids with which we share much in common, such as Neanderthal [7], had phonetic capacity far in excess of that ascribed to them by some [21]. If the phylogenetic reduction of oral cavity length is really as important as has been suggested [26], we would speculate that the use of laryngeally constricted postures/sounds might have played an even more central role in modulating vowel qualities in Neanderthal phonologies than in those of humans today.

We have ultimately argued that the laryngeal vocal tract is the locus of phonetic exploration and that it would seem that the sequence of phonetic acquisition takes advantage of this initially predisposed constricted posture of the larynx and on its subsequent unlocking. The overall process of phonetic acquisition is thus interacting with an already-in-place sequence of events that unfold during post-natal development and, furthermore, might also have placed some selective pressure on the shape and developmental sequence of the vocal tract itself.

5. REFERENCES

[1] Benner, A. 2009. *Production and Perception of Laryngeal Constriction in the Early Vocalizations of*

Bai and English Infants. PhD dissertation, University of Victoria.
[2] Benner, A., Grenon, I., Esling, J. H. 2007. Infants' phonetic acquisition of voice quality parameters in the first year of life. *Proc. 16th ICPhS* Saarbrücken, 2073–2076.
[3] Bettany, L. 2004. *Range Exploration of Pitch and Phonation in the First Six Months of Life*. MA thesis, University of Victoria.
[4] Boë, L.-J., Badin, P., Ménard, L., Captier, G., Davis, B., MacNeilage, P., Sawallis, T. R., Schwartz, J.-L. 2013. Anatomy and control of the developing human vocal tract: A response to Lieberman. *J. Phonetics* 41(5), 379–392.
[5] Burling, R. 2005. *The Talking Ape: How Language Evolved*. Oxford: OUP.
[6] Catford, J. C. 2001. *A Practical Introduction to Phonetics*, 2nd ed. Oxford: OUP.
[7] Dediu, D., Levinson, S. C. (2013). On the antiquity of language: The reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in Psychology* 4. doi:10.3389/fpsyg.2013.00397
[8] Edmondson, J. A., Esling, J. H. 2006. The valves of the throat and their functioning in tone, vocal register, and stress: Laryngoscopic case studies. *Phonology* 23, 157–191.
[9] Esling, J. H. 1996. Pharyngeal consonants and the aryepiglottic sphincter. *JIPA* 26, 65–88.
[10] Esling, J. H. 2005. There are no back vowels: The Laryngeal Articulator Model. *Canadian Journal of Linguistics* 50, 13–44.
[11] Esling, J. H. 2010. Phonetic notation. In: Hardcastle, W. J., Laver, J., Gibbon, F. E. (eds), *The Handbook of Phonetic Sciences,* 2nd ed. Oxford: Wiley-Blackwell, 678–702.
[12] Esling, J. H. 2012. The articulatory function of the larynx and the origins of speech. *Proc. Annual Meeting Berkeley Linguistics Society* 38, University of California, Berkeley. (via *eLanguage* of the LSA)
[13] Esling, J. H., Benner, A., Bettany, L., Zeroual, C. 2004. Le contrôle articulatoire phonétique dans le prébabillage. *Actes des XXVes Journées d'Étude sur la Parole*, 205–208, Fès, Maroc: AFCP.
[14] Esling, J. H., Harris, J. G. 2005. States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In: Hardcastle, W. J., Beck, J. M. (eds), *A Figure of Speech: A Festschrift for John Laver*. Mahwah, NJ: Lawrence Erlbaum, 347–383.
[15] Falk, D. 1975. Comparative anatomy of the larynx in man and the chimpanzee: Implications for language in Neanderthal. *Am J Phys Anthropol.* 43(1), 123–132.
[16] Gauffin, J. 1977. Mechanisms of larynx tube constriction. *Phonetica* 34, 307–309.
[17] Gick, B., Stavness, I. 2013. Modularizing speech. *Frontiers in Cognitive Science* 4, 1–3.
[18] Grenon, I., Benner, A., Esling, J. H. 2007. Language-specific phonetic production patterns in the first year of life. *Proc. 16th ICPhS* Saarbrücken, 1561–1564.
[19] Honda, K. 1995. Laryngeal and extra-laryngeal mechanisms of F0 control. In: Bell-Berti, F., Raphael, L. J. (eds), *Producing Speech: Contemporary Issues: For Katherine Safford Harris*. Woodbury, NY: American Institute of Physics, 215–232.
[20] Lieberman, P., Crelin, E. S. 1971. On the speech of Neanderthal Man. *Linguistic Inquiry* 2(2), 203–222.
[21] Lieberman, P., McCarthy, R. C. 2015. The evolution of speech and language. In: Henke, W., Tattersall, I. (eds), *Handbook of Paleoanthropology*. Berlin: Springer, 1–41. doi: 10.1007/978-3-642-27800-6_79-1
[22] MacNeilage, P. F. 1998. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21, 499–511.
[23] Moisik, S. R. 2013. *The Epilarynx in Speech*. PhD dissertation, University of Victoria.
[24] Newman, J. D. 1985. The infant cry of primates. In: Lester, B. M., Boukydis, C. F. Z. (eds), *Infant Crying*. New York: Plenum Press, 307–323.
[25] Nishimura, T. 2003. Comparative morphology of the hyo-laryngeal complex in anthropoids: Two steps in the evolution of the descent of the larynx. *Primates* 44(1), 41–49.
[26] Nishimura, T., Mikami, A., Suzuki, J., Matsuzawa, T. 2006. Descent of the hyoid in chimpanzees: Evolution of face flattening and speech. *J. Human Evolution* 51(3), 244–254.
[27] Stevens, K. N. 1989. On the quantal nature of speech. *J. Phonetics* 17, 3–45.
[28] Takemoto, H. 2008. Morphological analyses and 3D modeling of the tongue musculature of the chimpanzee (Pan troglodytes). *American Journal of Primatology* 70(10), 966–975.
[29] Titze, I. R. 2008. Nonlinear source–filter coupling in phonation: Theory. *JASA* 123(5), 2733.

ANATOMICAL BIASING AND CLICKS: PRELIMINARY BIOMECHANICAL MODELLING

Scott R. Moisik, Dan Dediu

The Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
scott.moisik@mpi.nl; dan.dediu@mpi.nl

ABSTRACT

It has been observed by several researchers that the Khoisan palate tends to lack a prominent alveolar ridge. A preliminary biomechanical model of click production was created to examine if these sounds might be subject to an anatomical bias associated with alveolar ridge size. Results suggest the bias is plausible, taking the form of decreased articulatory effort and improved volume change characteristics, however, further modelling and experimental research is required to solidify the claim.

Keywords: clicks, hard palate, alveolar ridge, anatomical biasing.

1. INTRODUCTION

This paper examines the production of clicks in the context of a three-dimensional biomechanical simulation. Specifically, we ask whether differences in the shape of the palate might influence certain aspects of click production, such as the muscular effort/articulatory ease, e.g. [9] and [13], or the dynamics of lingual cavity rarefaction. This work is situated within the larger context of research that seeks to address the question of whether variation in human vocal tract anatomy and physiology constitutes a systematic bias or pressure on speech sound systems. Such biases, while interesting at the level of individual variation, might also show localized patterns corresponding to wider populations of speakers sharing certain vocal tract traits.

It is an undeniable fact that human populations vary in certain systematic ways in their anatomy and physiology. This is true at both micro- and macroscopic levels, and advances in genetics will continue to elucidate the extent of these patterns of variation across populations. Early in the development of modern phonetic and phonological science, several proposals (e.g. [24] and [2]) were made which held that some of the diversity observed in speech sound systems around the globe might be owing to systematic variation observed in the anatomy and physiology of the speakers of language, in addition to the other factors driving language change and diversification. These ideas were hastily dismissed as implausible, on the grounds that any human being can learn any human language.

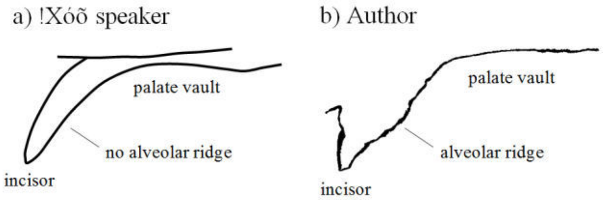
It is an incontrovertible fact that normal variation of the human vocal tract does not preclude an individual from acquiring any spoken language. However, the hypothesis that human vocal tract morphology exerts a bias on the way we speak seems plausible, and the possibility that such biases might have expressions at the level of populations of speakers has never been satisfactorily ruled out. It also seems to have resulted in the unfortunate side-effect that details of vocal tract shape are rarely if ever correlated to production variables in phonetic research. A relatively recent return to the question of whether normal vocal tract variation can indeed exert such biases reflects the unresolved nature of the problem. Many examples exist for such research examining the individual level (e.g. [25], [3], and [18]), and these are laden with implications for impacts at broader levels, with some researchers even suggesting it may be a driver of change of certain aspects of entire phonological systems (e.g. [1], [5], and [17]).

1.1. Why examine click production?

In the present study, we focus on the case of clicks. Clicks merit investigation because of their incredible rarity as phonemes, a fact which suggests there are biases against the phonological incorporation of these sounds. They are primarily associated with the so-called Khoisan languages (actually a group of language families, including Kx'a, San, and Tuu, which bear some family resemblance, and the isolates Hadza and Sandawe). They are also found in several Nguni Bantu languages (including Zulu, Xhosa, Ndebele, Swazi, and Sotho) and Dahalo, a Southern Cushitic language, all of which have evidently borrowed clicks through generations of extensive contact with various Khoisan languages [20].

Our inspiration for the present study comes from observations by Engstrand [6] (also [20], p. 4) and Demolin (p.c.) that clicks may be subject to a production bias grounded in the morphology of the palate. The ultimate source for this idea comes from Traill [21] (p. 101-102), who remarks in his dissertation (on the subject of !Xóõ, a language of the Khoisan group) that one cannot use the term *alveolar* to describe post-dental clicks in !Xóõ since four of his five subjects “do not have an alveolar ridge” (p. 101). One of these palates is reproduced in Fig. 1 along with a comparison to the palate of author SRM, which exhibits a sizeable alveolar ridge.

Figure 1: Mid-sagittal palate profiles: (a) an example of a !Xóõ speaker’s palate (retracing of Fig. 24 from Traill [21], p. 107) and (b) the palate of author SRM.

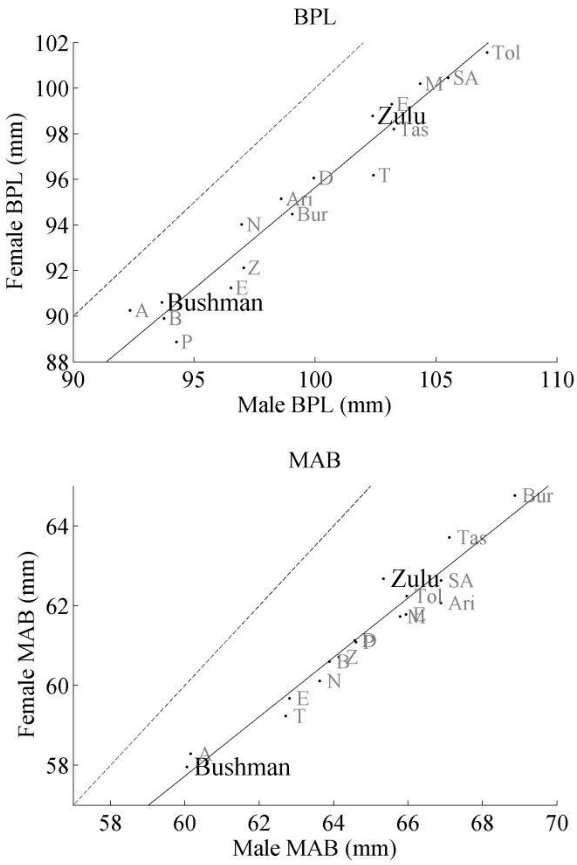


While such variation could easily be owing to Traill’s limited sample of !Xóõ palates (although Traill notes that the pattern holds for the San in general, citing [22]), it is well established that other members of the Khoisan group show uniformity of head and palate morphology that distinguishes these groups from other nearby non-Khoisan populations. For example, [23] compares palatal measures made on plaster dental casts of Central Kalahari Bushmen (a sample comprised of individuals from the !Kung, Auen, Naron, Dukwe, and Heikum tribes), Vassekela Bushmen (originating from Angola); and Herero-speaking individuals, mainly Himbas, for comparison. Note that the former two groups (the Bushmen) consist of speakers of Khoisan click languages, but Herero (a Bantu language) lacks clicks. Sample sizes in this study are large (minimum of 76 and maximum of 158). The Bushmen groups generally have narrower and shallower palates, and the anterior flatness (i.e. lack of a prominent alveolar ridge reflected by highest scores for palate height in the canine region) is confirmed. The Vassekela Bushmen are intermediate, but classified with the Himbas as having a “shelved” palate: low at the front but suddenly increasing in height towards the back. The Bushmen palates were not necessarily shorter than those of the Himba.

Similar work [26] compares 110 male !Kung San (who speak a Khoisan language of Namibia) with a group of 138 males from Kenya and Uganda (containing both Bantu- and Nilotic-speaking individuals). This study demonstrates that the !Kung San palate is shorter, narrower, and shallower and characterized by a smooth, concave profile. Note that the authors of [26] do not provide a detailed listing of the specific languages spoken by the non-Khoisan group, i.e. the Bantu and Nilotic speakers. However, it is stated that most of the Bantu-speaking individuals are from the Taita Hills, and the language of this area, Taita/Dabida, lacks clicks; and clicks are not found in Nilotic languages.

Craniometric data [8] show that Bushmen (Khoisan) palates (for males or females) tend to be smaller in comparison to many other populations (Fig. 2). Note that Zulus, whose language has clicks, fall towards the upper end of these variables.

Figure 2: Basion-prosthion length (BPL; proxy for palate length) and maxillo-alveolar breadth (MAB; proxy for palate width). Data from [8]. A = Andaman, Ari = Arikara, B = Berg, Bur = Buriat, D = Dogon (Mali), E = Egyptian, E = Eskimo, M = Mokapu, N = Norse, P = Peru, SA = South Australian, T = Teita (Kenya), Tas = Tasmanian, Tol = Tolai, Z = Zalavar. Dashed line = hypothetical 1:1 sexual dimorphism; Solid line = regression line.



In short, it seems that the Khoisan palate is distinguishable from palates of other groups, and that the trend of a lack of a prominent alveolar ridge detected in Traill’s x-rays may indeed be representative of the Khoisan group, although gene flow with neighbouring groups and the resultant differentiation of palate shape (e.g. as reflected in the Vassekela) is a possibility.

1.2. Palate morphology and clicks: Hypotheses

Briefly, to produce a (lingual) click, the tongue must first form an enclosed space between the anterior occlusion (which defines the click's place of articulation) and the velar-uvular region of the palate. Clicks do not typically require the tongue be flush against the palate, and, in fact, there is very often a central gap, as observed in x-ray ([21] and [10]), static palatography ([21] and [19]), and ultrasound (e.g. [12]) studies. The next step is to generate the velaric ingressive airstream, which depends on rarefaction of the air driven by localized lowering of the tongue body (the exact location of which is dependent upon click place of articulation). Finally, the oral seal is suddenly broken by the rapid release of the anterior occlusion, and the pressure differential created through rarefaction yields a transient acoustic signal audible as a click.

Our goal was to probe into the possibility that palatal morphology has consequences for click production and that this, in turn, might speak to a production bias which has led to the establishment and maintenance of clicks as speech sounds. The general question we ask is: what effect, if any, does palate shape have on the production of clicks?

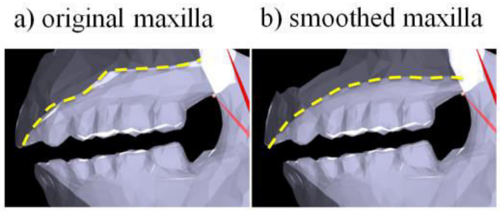
To address this question, we narrow our focus on the biomechanics of click production, and, on the alveolar ridge, which was identified as an important factor by other researchers. (Palatal dimensions may also be important, but in this preliminary modelling, these factors were not explored.) Given this focus, we suggest the following hypotheses regarding alveolar ridge shape and click production: (1) a smooth palatal profile requires less articulatory effort to form click stricture since the anterior tongue does not need to deform as much to form the lingual seal; (2) a smooth palate provides better volume change characteristics (presumably for achieving efficient aero-acoustic effects in click production, although this was not modelled).

To test these hypotheses, we assume that total muscle force is a good proxy for articulatory effort (following [9] and [13]). We also constrain our attention to the production of clicks which involve contact between the tongue tip/blade and the anterior palate, as these clicks are most relevant to hypothesis (1). Our simulations are place-abstract, but they most closely resemble dental clicks.

2. METHODOLOGY: CLICK SIMULATION IN ARTISYNTH

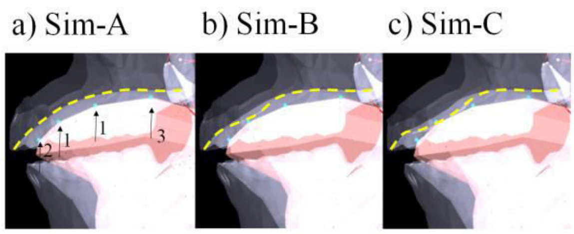
The biomechanical simulation of click production was created using the ArtiSynth biomechanical modelling toolkit (www.artisynth.org; [11]). This model is based on the 3-D finite-element (FE) tongue integrated with rigid-body skeletal structure for the maxilla and mandible as originally presented in [4] (and used in several subsequent studies; see [14], [16], and [17]).

Figure 3: Geometry (a) before and (b) after maxillary smoothing in the region of the alveolar ridge (midsagittal profile). The yellow dashed line highlights the contour of the mesh for comparison.



Alveolar ridge shape was systematically manipulated to simulate its effects on click production. To do this, it was first necessary to smooth the original maxillary geometry, which features a prominent alveolar ridge. Smoothing was accomplished manually using tools in Blender (www.blender.org) to deform the anterior palatal geometry such that the alveolar ridge convexity was entirely removed. Results of this process are illustrated in Fig. 3.

Figure 4: Mesh warping to control alveolar ridge size. Three simulation conditions (a) Sim-A, no warping, "no ridge"; (b) Sim-B, mild warping, "small ridge"; (c) heavy warping, "big ridge". The yellow dashed line highlights the change in profile. Arrows show longitudinal locations of inverse-simulation nodes (see text below).



Next, to experimentally manipulate the shape of the alveolar ridge, a spherical warping field was used. This field radially displaces subjected mesh vertices within a limited radius of the origin of the warping field (which was placed approximately above the anterior nasal spine). The magnitude of the displacement is given by $d = (r - p)/r$, where p is the Euclidean distance between a given vertex and the warping origin, and r is the radius of the warping field. The different grades of warping used are shown in Fig. 4 (note that the warping in Sim-B is intermediate between Sim-A and Sim-C).

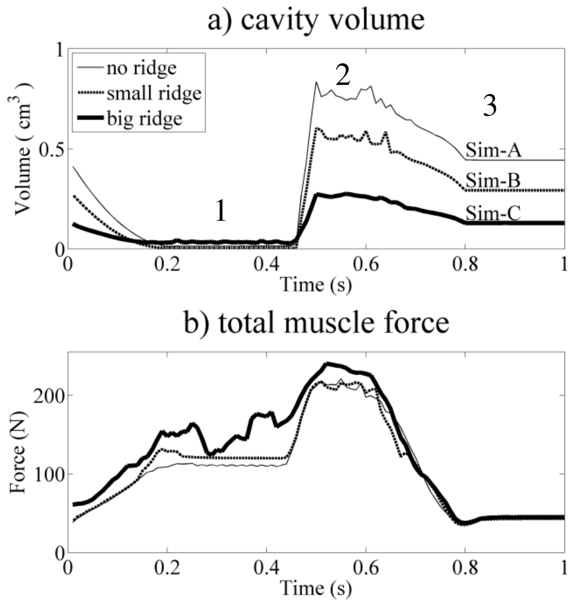
Finally, to simulate the dynamics of click production, ArtiSynth's inverse controller was used. This takes temporal targets of nodal locations of the geometry as input and outputs a parsimonious set of muscle activations which achieve these temporal targets within the limitations set by tissue contacts, inertia, and material properties. Inverse targets were associated with FE nodes at longitudinal locations shown in Fig. 4a (blue circles or arrows), each of which had one midline and two lateral nodes. A rudimentary, somewhat idealized and place-neutral lingual click was defined as follows: first, all inverse targets were positioned at a short distance beyond the projection of each target's corresponding FE node onto the nearest face of the maxilla mesh along the line of projection (thus, in each simulation, constriction is relative to maxilla shape); then, the midline nodes at the positions indicated by the two arrow-1s (Fig. 4a) were displaced to a position below their resting state positions (this simulated rarefaction); next, all targets at arrow-2 (Fig. 4a) were displaced to their resting state (simulating release of the front closure); finally targets at arrow-3 were returned to resting state (simulating release of the back closure). Note that no attempt was made to simulate the initial presence of an enclosed airspace during the establishment of palatal contact.

Three 1 second simulations were run which correspond with the geometries in Fig. 4. Total muscle force was observed along with the volume in the region of lingual rarefaction (arrows 1 and 2).

3. RESULTS AND DISCUSSION

Fig. 5 shows results for volume change and total muscle force. Overall, the effect of having a larger alveolar ridge, given the same relative palatal contact requirements and the same absolute lingual resting/return state, is to reduce the rate and amount of volume gain during release of the front closure (Fig. 5a, at 0.5 s) and to increase the articulatory effort in producing and maintaining closure whilst enlarging the air space. Also note that volume did not go to zero for the "big ridge" condition (phase 1, Sim-C), which indicates that this condition makes it harder for the model to establish full contact between the tongue and the palate.

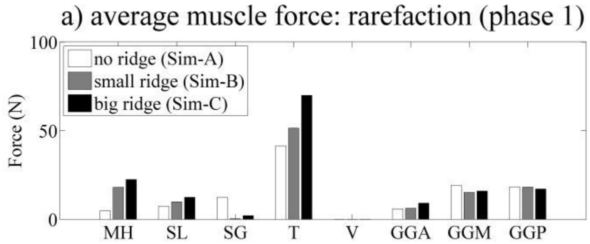
Figure 5: Lingual cavity volume (a) and total muscle force, smoothed with a moving average filter (b) for the three simulation conditions (see Fig. 4). Phases: 1 = rarefaction; 2 = release of front closure; 3 = return to resting state.

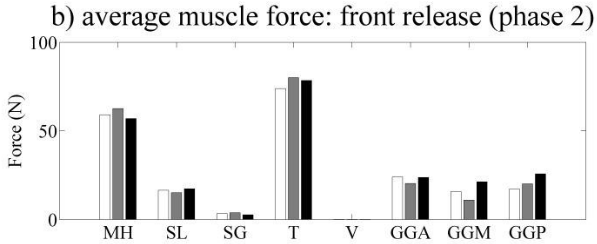


When examining specific muscle contributions (see Fig. 6), it is apparent that, during the rarefaction phase (Fig. 5a, 1; Fig. 6a), mylohyoid (MH) and transversus muscle force output increase with alveolar ridge size, followed by the superior longitudinals and then, somewhat less so, by genioglossus anterior (GGA). The styloglossus and genioglossus medial (GGM) muscle outputs are actually higher for the "no ridge" condition (Sim-A).

The rather large values for the transversus muscles during phase-1 (Fig. 1a) can be associated with intrinsic lingual shaping to form and maintain contact against the palate. In the simulation, the rarefaction is probably driven by the GGM fibres; the verticalis might also play a role in real productions but it is inactive here. Relative to those muscles responsible for elevating the tongue against the palate, the activity of GGM seems low. Furthermore, during front release, MH and transversus are still very high which, in an effort to maintain the posterior closure, is possibly occurring to balance the forces working to release the front closure (mainly the GGA). Validation of these muscle activation patterns would be difficult to achieve with electromyography (and the authors are unaware of any such study for click articulation). Refinements to the geometry of the tongue might change the patterns substantially.

Figure 6: Average muscle force for the rarefaction (a, phase 1) and release of front closure (b, phase 2) phases corresponding to 1 and 2 in Fig. 5a. Muscles: MH = mylohyoid; SL = superior longitudinal; SG = styloglossus; T = transversus; V = verticalis; GGA = genioglossus anterior; GGM = genioglossus medial; GGP = genioglossus posterior.





The simulation exhibited some peculiarities. First, as is evident in Fig. 5, there is some noisiness, which is directly attributable to the interaction of collision mechanics and the inverse solver in ArtiSynth: if the inverse targets go beyond a site of collision, the inverse solver will continuously oscillate through various solutions. To minimize this, targets where placed as close as possible to the palate but still slightly above so as to ensure strong contact. Also, it was apparent that the tongue FEM discretization was not fine enough to achieve an anterior lingual deformation during rarefaction (Fig 5a, phase 1) sufficient to produce a gradual expansion of the volume (from phase 1 to 2). This may have also been the cause of the somewhat unexpectedly large muscle forces occurring at stage 2. Also note that, while in reality it may be that the negative pressure generated from rarefaction requires heightened muscle forces during this phase, no fluid-structure interaction was simulated, so this cannot be the cause of the increased force at this point. Finally, no attempt has been made to model the active contribution of the soft palate in the formation of the velar closure in click production. These aspects need to be resolved in future refinements to the model.

With these considerations of the limitations of this preliminary ArtiSynth model of click production in mind, the results are consistent with the hypotheses introduced in §1.2: (1) more muscle force is required to form click stricture with a larger alveolar ridge, and (2) all things being equal, the smoother the palate, the more rapid and larger the volume change. We suspect that greater articulatory effort (estimated through total muscle force) will have a negative bias on click appearance and maintenance at the diachronic scale. Larger and faster volume change ought to produce acoustically stronger click bursts with better transient properties. It also provides a wider range of volumes achievable depending on other factors, and this should increase the reliability of click production (cf. [3]). Finally, incomplete lingual-palatal contact in Sim-C could indicate that the alveolar ridge inhibits efficient lingual sealing, although finer FE discretization needs to be tested. The viscosity of saliva on the tongue and palate may also influence click biomechanics and consideration of these forces could be incorporated into future models.

This modelling supports the notion that alveolar ridge shape may be a source of biasing on clicks, but one that is weak at best. The borrowing of clicks by non-Khoisan groups with possibly quite different palate size (e.g. see Zulu, Fig. 2) and shape support this interpretation of a weak bias. Furthermore, clicks are a common paralinguistic sound, and they are often spontaneously produced by children. On this last point, however, it is possible that children, regardless of alveolar ridge size, benefit in click production from having overall smaller palate dimensions, not unlike the Khoisan (Fig. 2). As noted, palate dimensions were not considered here, but one can imagine how a narrow palate might facilitate click seal formation (although tongue size is relevant, too). Palate size might also influence the amount of pressure exerted by the tongue on the teeth [15].

In this preliminary work, the model abstracts away from place of articulation. However, place is likely important, and the details of muscle forces and volume change characteristics are very likely to be a function a click place of articulation. In particular, given the relative rarity of palatal clicks and their resistance to borrowing (found only in Yeyi [7] outside of the Khoisan group), these may be most strongly subject to a bias. The direction of lingual motion in such clicks is different and could be a source of differential articulatory efficiency determined by palate shape. We intend to explore different places of click articulation in subsequent modelling work.

4. REFERENCES

[1] Allott, R. 1994. Motor theory of language: The diversity of languages. In J. Wind, A. Jonker, R. Allott, & L. Rolfe (Eds.), *Studies in Language Origins*

(Vol. 3, pp. 125–160). Amsterdam & Philadelphia: John Benjamins Publishing Comany.

[2] Brosnahan, L. F. 1961. *The Sounds of Language: An Inquiry into the Role of Genetic Factors in the Development of Sound Systems*. Cambridge: W. Heffer and Sons.

[3] Brunner, J., Fuchs, S., Perrier, P. 2009. On the relationship between palate shape and articulatory behavior. *JASA* 125(6), 3936–3949.

[4] Buchaillard, S., Perrier, P., Payan, Y. 2009. A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning. *JASA* 126(4), 2033–2051.

[5] Dediu, D. 2011. Are languages really independent from genes? If not, what would a genetic bias affecting language diversity look like? *Hum. Biol.* 83(2), 279–296.

[6] Engstrand, O. 1997. Why are clicks so elusive? *PHONUM (Umeå University)* 4, 191-194.

[7] Fulop, S. A., Ladefoged, P., Liu, F., Vossen, R. 2003. Yeyi clicks: Acoustic description and analysis. *Phonetica* 60(4), 231–260.

[8] Howells, W. W. 1973. Cranial variation in Man: A study by multivariate analysis of patterns of difference among recent human populations. *Papers of the Peabody Museum of Archaeology and Ethnology* 67, 1–259.

[9] Kirchner, R. 1998. An effort-based approach to lenition. Diss. UCLA.

[10] Ladefoged, P., Traill, A. 1984. Linguistic phonetic descriptions of clicks. *Language* 60(1), 1–20.

[11] Lloyd, J. E., Stavness, I., Fels, S. S. 2012. ArtiSynth: A fast interactive biomechanical modeling toolkit combining multibody and finite element simulation. In Y. Payan (Ed.), Soft tissue biomechanical modeling for computer assisted surgery (pp. 355–394). Berlin, Germany: Springer.

[12] Miller, A., Scott, A., Sands, B. E., Shah, S. 2009. Rarefaction gestures and coarticulation in mangetti dune! xung clicks. *Proc. of Interspeech*, Brighton, 2279–2282.

[13] Napoli, D. J., Sanders, N., Wright, R. 2014. On the linguistic effects of articulatory ease, with a focus on sign languages. *Language* 90(2), 424–456.

[14] Nazari, M. A., Perrier, P., Chabanas, M., Payan, Y. 2011. Shaping by stiffening: a modeling study for lips. *Motor Control* 15(1), 141–168.

[15] Proffit, W. R., McGlone, R. E., Barrett, M. J. 1975. Lip and Tongue Pressures Related to Dental Arch and Oral Cavity Size in Australian Aborigines. *J. Dent. Res.* 54(6), 1161–1172.

[16] Stavness, I., Lloyd, J. E., Payan, Y., Fels, S. S. 2011. Coupled hard-soft tissue simulation with contact and constraints applied to jaw-tongue-hyoid dynamics. *Int. J. Num. Meth. Biomed. Eng.* 27(3), 367–390.

[17] Stavness, I., Nazari, M. A., Perrier, P., Demolin, D., Payan, Y. 2013. A biomechanical modeling study of the effects of the orbicularis oris muscle and jaw posture on lip shape. *JSLHR* 56(3), 878–890.

[18] Stone, M., Rizk, S., Woo, J., Murano, E. Z., Chen, H., Prince, J. L. 2013. Frequency of apical and laminal /s/ in normal and post-glossectomy patients. *JMSLP* 20(4), 106–111.

[19] Thomas-Vilakati, K. D. 1999. Coproduction and coarticulation in IsiZulu clicks. Diss. UCLA.

[20] Traunmüller, H. 2003. Clicks and the idea of a human protolanguage. *PHONUM (Umeå University)* 9, 1–4.

[21] Traill, A. 1985. Phonetic and phonological Studies of !Xóõ Bushman. In *Quellen zur Khoisan-Forschung* (Vol. 1). Hamburg: Helmut Buske Verlag.

[22] Van Reenen, J. F. 1964. Dentition, jaws and palate of the Kalahari Bushman. *J. Dent. Assoc. S. Africa* 19, 1–37.

[23] Van Reenen, J. F., Allen, D. W. 1987. The palatal vault of the Bushman (San), Vassekela and Himba. *J. Dent. Assoc. S. Africa* 42(8), 489–492.

[24] Vendryès, J. 1902. Some thoughts on sound laws. In A. R. Keiler (Ed.), *A Reader in Historical and Comparative Linguistics* (pp. 109–120). New York: Holt, Rinehart and Winston.

[25] Weirich, M., Fuchs, S. 2011. Vocal tract morphology can influence speaker specific realisations of phonemic contrasts. *Proc. ISSP* 251–258.

[26] Winkler, E. M., Kirchengast, S. 1993. Metric characters of the hard palate and their cephalometric correlations in Namibian !Kung San and Kenyan tribes. *Hum. Biol.* 65(1), 139–150.

EXPLORING POTENTIAL CLIMATIC EFFECTS ON THE EVOLUTION OF HUMAN SOUND SYSTEMS

Seán G. Roberts[1], Caleb Everett[2] & Damián Blasi[3]

1 Language and Cognition Department, Max Planck Institute for Psycholinguistics; 2 University of Miami; 3 Max Planck Institute for Mathematics in the Sciences, Max Planck Institute for Evolutionary Anthropology
sean.roberts@mpi.nl, caleb@miami.edu, damian.blasi@mis.mpg.de

ABSTRACT

We suggest that it is now possible to conduct research on a topic which might be called *evolutionary geophonetics*. The main question is how the climate influences the evolution of language. This involves biological adaptations to the climate that may affect biases in production and perception; cultural evolutionary adaptations of the sounds of a language to climatic conditions; and influences of the climate on language diversity and contact. We discuss these ideas with special reference to a recent hypothesis that lexical tone is not adaptive in dry climates [17].

Keywords: There is space for up to five self-selected keywords (maximally two lines).

1. INTRODUCTION

In theory, there are three ways in which climate could affect language. First, linguistic articulators may be biologically adapted to aspects of climate. These adaptations could shape the possible space of language sounds. Secondly, the climate can affect aspects of production and perception which may cause aspects of language to adapt via cultural evolution. For example, the vocal folds are affected by the inhalation of dry air in a way that affects phonation. Also, in spoken languages, the air is the interface between production and perception, and so a possible source of noise or bias in cultural transmission. Finally, the climate may have indirect affects on cultural evolution by influencing population migration and contact which could drive innovation and divergence.

The main focus of evolutionary linguistics has been to identify universal properties of languages and link them to conditions of genetics or culture that humans have in common. However, some studies have also considered how idiosyncratic aspects of a linguistic community might affect the development of its language, such as differences in demography [4,23] or genetic biases [6,10]. In a similar way, we suggest that it is possible to research the differences in language based on climatic differences. In this paper we delineate the three ways in which climate could possibly impact the evolution of language. We sketch out the various questions and problems of such a line of research, and offer a clear methodological path for this line of research to be pursued constructively. As a case study, we use our recent study of the idea that the inhalation of dry air makes the precise control of tone difficult, leading, via cultural evolution, to fewer languages using lexical tone in dry environments [17].

2. POTENTIAL EFFECTS

2.1. Climate and biological evolution

Many animal communication systems show adaptation to the environments in which they are used. Animal signals adapt to environmental noise and obstructions to signals such as plant cover (e.g. [28]; see [39]; though see [5]), and some studies have found these factors to be better determiners of song properties than climate [21]. However, temperature and humidity also affect acoustic absorption [3], meaning that dry, warm environments have greater absorption of high frequencies. Bird and bat signals adapt to higher absorption climates by using narrower bandwidth signals that carry further in these conditions, with bats adapting within climates over seasons as well as between climates [34] (though songs are learned to some extent, so there may be gene-culture co-evolution).

Insect chemical signals are also adapted to humidity and temperature, which affect evaporation and diffusion rates [1].

Biological adaptations to the climate may also have knock-on effects for language. For example, the morphology of the nasal cavity has evolved in different human populations so that those in drier, colder climates are higher and narrower which increases the contact between air and nasal wall, helping to humidify the inhaled air [30]. These adaptations could have small effects on nasal sounds used in speech production [16]. Interestingly, recent work has also demonstrated altitude-based effects of the formants associated with nasal phonemes [31].

More generally, we also note that long-term, pre-historic changes to climate have been linked with more general adaptations such as bipedalism, which may have allowed larger human group size [22]. Both of these aspects have been suggested as pre-adaptations for language [22,40], suggesting a link between differences in climate and differences between species.

2.2. Climate and cultural evolution

The basic principle behind studies of cultural evolution is that a selective pressure on communication can transform the structures of a language over time. Based on a similar line of argument to the ecological adaptiveness of animal signalling, [19] suggest that sonorous speech sounds are better at carrying longer distances, and so would be more adaptive in environments where plant cover was dense, and hence warmer climates (see [8] for a direct test of plant coverage using spatial regression). This combines with assumptions about aspects of culture such as communities in warmer climates spending more time outdoors and therefore also communicate over relatively large distances. This theory does not involve a direct influence on the climate on as in [17], but rather an interaction between the climate, the ecological environment and interactional norms that bring about a selective pressure (see also [15]).

Perhaps one of the reasons that the climate has not been more widely considered as a selective pressure on languages is the focus on language learning as the locus of language change. Acquisition has been conceptualised by some as primarily a cognitive task, and there is no theory that would predict substantial differences in formal learning systems nor neural functioning based on climate. However, if we see the locus of language change as the production and perception of individual utterances [9], then the interface between the physical articulators and the medium of communication (the air) becomes more salient.

Another reason that the influence of climate might be doubted is the known role of social factors on language change. Languages die, survive or change based on historical events, power, politics and socioeconomic factors. Given that the effect of climate on language should be subtle (it's certainly not impossible to speak a tone language in dry air) and take a long time to propagate, it's possible that these effects could be masked by the more powerful social forces. However, this is an empirical question, and large-scale cross-linguistic databases make it feasible to detect subtle influences.

2.3. Climate and diversity

There has also been some implication of indirect influence of the climate on linguistic diversity. Essentially, the climate can affect the 'carrying capacity' of the environment, affecting demographics of speakers. Nettle [29] argues that certain climates and ecologies foster certain kinds of social interaction between linguistic groups. A high carrying capacity leads to demand for material wealth, so linguistic groups invest in each other through learning each others' languages, causing linguistic diversity to be an asset. Nettle finds correlations between linguistic diversity and climatic factors such as temperature and mean growing season.

Linguistic change is also brought about through migration (leading to isolation) and contact. The climate can influence migration patterns (e.g. extreme changes in climate can force groups to move) or influence where contact is likely to happen. [20] estimate the timing of divergence of languages in the Uralic family using Bayesian phylogenetic techniques, and compare this to the changes in climate. They argue that changes in climate align with linguistic divergence. One suggested explanation is that a rise in temperature leads to a rise in population size [36], which makes migration more likely (though see [32] for an argument that innovation drives migration). Similarly, a decrease in temperature can decrease the population, leading to conservatism.

2.4. Interactions between climate, biology and language

Aspects of climate, biological evolution and culture may not be independent from one another. For example, an adverse affect of climate on phonation may be adapted to biologically (e.g. increasing saliva production, a longer, narrower nasal cavity in response to drier air) or culturally (e.g. the cultural practices of breathing in particular ways to avoid desiccation of the mouth and larynx). This may cause two problems. First, climatic differences may be neutralised by biological adaptations, meaning that there is no difference in the effect of climate on production. Secondly, variation in genetics or morphology could mean that climate may not affect production in the same way in all populations. Controlling for this is difficult, but one solution is to gather cross-cultural data, combining knowledge from geography, genetics, linguistics and anthropology.

3. DEMONSTRATING CAUSAL EFFECTS

A study in evolutionary geophonetics, focussing on cultural evolution, would ideally proceed in the following way. First, evidence is obtained that a change in a certain property of the climate causes a change in production or perception. For example, evidence that a property of climate affects the articulators of language that leads to a difference in production, such as the inhalation of dry air causing changes to the vibration of the vocal folds that affect acoustic properties of phonation. This is not always straightforward, since measurements can involve invasive methods (cf. [25]). Alternatively, an effect could be demonstrated on perceptual systems (e.g. hearing being affected by temperature or humidity, [27]) or on the way sound is carried in different climates [2]. We also note that similar predictions can be made for languages in other mediums, for instance [33] discusses the impact of temperature on sign language in the Arctic.

Crucially, one must be able to demonstrate that the effects on one aspect of interaction lead to differences in the other, for example that the changes in phonation caused by dry air lead to differences in perception. Usually, this difference will involve a difference in a specific aspect of production, rather than an effect across the board (which may make predictions more difficult).

Once a physical link is proposed, a prediction can be made about the way in which languages will change in different climatic environments. This involves a prediction of how individual interactions will be affected, and also how those interactions will accumulate into wider change. In general, the prediction will be based on cultural evolutionary principles: the climate provides a selective pressure which causes differential rates of successful production and perception for particular linguistic aspects. Predictions may not be straightforward to make, and may involve computational models of both articulation (e.g. [26]) and cultural evolution (e.g. [35]). This leads to a concrete prediction of how a given property of language will co vary with a property of the environment.

This can then be tested in several ways. A synchronic pattern can be identified in current languages. This is not straightforward either due to the non-independence of languages and other statistical concerns. However, it provides evidence that the current state of linguistic distributions is compatible with the prediction

Diachronic evidence is also obtainable. This could be done by a case-study of the divergence of two languages with the expected differences. More large-scale studies are also possible by reconstructing linguistic and climatic history. Linguistic history can be estimated from current data using phylogenetic techniques [14]. Similarly, climatic history can be reconstructed from current sources such as sediment or pollen. One can then test whether a change in the environment coincides with the predicted change in the linguistic property.

We also note that the general principle of linguistic change due to climatic influence could be demonstrated through experimental techniques such as iterated communication games [37]. Such techniques may, in and of themselves, be used to demonstrate reasonable causal interpretations of the associated distributional data.

Following these steps, one may arrive at a demonstration of a causal link between climate and language. However, we note that not all studies are immediately possible, and some unfeasible without considerable effort. Different kinds of evidence can be collated piecemeal to provide a robust argument.

3.1. A case study: Tone and Humidity

[17] review the literature on the effects of inhaling dry air on the larynx and vocal folds. Dry vocal folds increase phonation threshold pressure and perceived phonation effort, and create a signal with higher rates of jitter and shimmer. In short, dry vocal folds make it harder to precisely control pitch (though there is not yet direct evidence that it influences production in a perceivable way). This is a more direct link between climate and language than suggested by previous studies.

All languages use pitch contrasts for various purposes, often pragmatic. Also, tonal contrasts are often not simply pitch-based but rely on other factors such as laryngealization. Additionally, F0 modulation can be as extreme in non-tonal languages as in tonal ones. However, it is still reasonable to assume that tonal languages, particularly those with complex tone, require that a generally higher burden be placed on the maintenance of precise pitch patterns in order to contrast meaning (we note that this may also be empirically testable).

Assuming that this puts a selective pressure on individual utterances, which is amplified by cultural evolution, leads to a prediction that languages in dry areas will not use lexical tone. While humidity is predicted to affect phonetic production, rather than phonology, we predict that, over time, biases in phonetic production affect changes to the phonological system.

[17] use a database of over 3,700 languages [11] to demonstrate a synchronic pattern: languages with complex lexical tone are rarer in areas of the world

with dry climates, and that this distributional tendency is not simply owed to genealogical or contact-based confounds.

The analysis was complicated by two factors. First, the languages were related historically, meaning that they did not constitute independent samples. Secondly, the prediction is a uni-directional implication: it suggests that complex tone should be rarer in cold climates, but makes no prediction about the distribution in warm climates. In this case, typical regression frameworks, which are suited to bi-directional implications, are not appropriate.

The solution was to use a Monte Carlo framework. Random samples of languages with complex and non-complex tones were taken and the distribution of humidity in each sample was compared. It was predicted that the distribution for complex tone languages would have a higher lower-quartile (more humid) than the non-complex languages (the mean of the two distributions could be similar at the same time as there being a 'gap' in the complex tone languages). This provided a direct way to test the prediction of a difference in low-humidity languages. The samples were balanced by selecting only one language from each language family, and by having the same number of languages in both the complex and non-complex samples. This addressed the first problem.

The study engendered discussion in various quarters, and attracted some scepticism. Some of this scepticism was, we suspect, the unfortunate byproduct of media reports suggesting e.g. a simple correlation between humidity and tonality—a position not propounded in our paper. Additionally, some of the scepticism resulted partly from the unfamiliar statistical methods used, and partly from the unusual claim that different languages may be subject to different evolutionary pressures rather than the more traditional bias towards studying effects that apply universally to speakers (e.g. processing, memory). There were also questions about why the effect should be seen specifically for lexical tone. We stress, however, that the link between humidity and lexical tone does not exclude the potential effects of humidity on other uses of pitch in language, such as clausal prosodic contours. Future work might explore this, but it is worth noting that the transmission or borrowing of lexical pitch and clausal pitch likely work quite differently. Criticisms of the suitability of the data on tone were more perspicacious. However, extant databases only allow us to test our hypothesis as it relates to major tonemic categories across languages. It is worth underscoring as well that, subsequent to the publication of our paper, no alternate hypotheses have been presented that explain the climatic-tonemic association we have uncovered. We emphasise that the hypothesis derives from an *a priori* prediction from known physical causes, and that it can be quantitatively tested.

It isn't yet known whether the link between tone and climate is truly supported by historical change, though the intra-family analyses offered in [17] do suggest that in four of the world's major language phyla historical patterns are congruent with the suggested causal effect. The expectation is that languages moving into dry areas will be less likely to gain tone contrasts in the first place, rather than dry air leading to loss of tone or humid air leading to the adoption of tone.

4. DISCUSSION

4.1. Implications for language change

Many studies of cultural evolution focus on cognitive selective pressures (e.g. processing, memory, frequency etc.), which are usually assumed to apply universally. The results in [17] suggest that some pressures may not be universal, but only apply in particular situations, for example in very dry contexts that influence vocal-tract physiology in particular ways. This adds to the literature on niche-specific cultural evolution, such as the effect of population size on morphological complexity, or demography on phoneme inventory.

4.2. Implications for language acquisition

An interesting question is whether the interaction between climate and production and perception will also affect acquisition, either learning a native language (e.g. from birth, L1 acquisition) or learning a language later in life (L2 acquisition).

With regards to L1 acquisition, difficulty in producing or perceiving sounds could lead to biased acquisition. Since languages must adapt to be learnable by children [7], this could also lead to language change. While [17] does not address this issue for tone, and while there is some evidence that pitch is an important cue in learning [18], we think this is not very likely. Children have differences in production due to developing articulators whose effects are likely to be much greater than those of climate, and which disappear with maturity. This hypothesis is also very difficult to test, given that a change in climate almost always brings with it a change in social factors, cultural contexts and linguistic phenomena that influence learning.

However, the case might be different for L2 acquisition. Adults find learning the phonetics and phonology of a new language challenging. L2 learning is also sensitive to psychological aspects such as confidence and motivation (e.g. [12]). So if sounds are harder to produce or perceive due to dry air, adult learners may find them harder to learn. In theory, this is testable by looking at learning performance over a range of climates. However, again, with a difference in climate comes a difference in culture, socioeconomic status, motivation and so on, which would complicate the answer.

4.3. Implications for linguistic typology

Theoretically, there are many other aspects of the sounds of a language that could be affected. Also, the sounds of a language can, in principle, have a knock-on effect on other parts of language like morphology or syntax. For example, [38] discuss the idea that lexical tone and phrase-level intonation compete for the same linguistic resource (pitch), and show that languages with lexical tone are more likely to develop additional grammatical means of distinguishing questions versus statements.

Other implications might be made for the semantics of temperature [23] and possible extensions into metaphor, but are not discussed here.

4.4. Implications for other aspects of culture

The general hypothesis offered in [17] might predict differences in music or singing styles. However, there are differences in the function between singing and language. Singing is often performative, while language is communicative. In this case, there may be less pressure on singing to adapt to the environment. In fact, performative pressures may act in opposition to pressures for simplicity and efficiency.

5. CONCLUSION

When we look at the world's languages, we see a lot of variation. Some aspects, like lexical tone, can seem completely alien to speakers of many European languages. Similarly, the variable stress patterns of languages like English can seem strange to speakers of other languages. However, rather than seeing these differences between languages as odd or due to chance, we suggest that languages are well adapted to the communicative needs of its speakers. In some cases, this can also mean adaptation to climate. Using the recent findings in [17], we have sketched out an initial heuristic approach to a nascent field of inquiry, one we have termed evolutionary geophonetics.

6. REFERENCES

[1] Alberts, A. (1992) Constraints on the Design of Chemical Communication Systems in Terrestrial Vertebrates. *The American Naturalist*, Vol. 139.

[2] Ancell, J. E., and Hardy, H. C. Effect of external sound fields on hearing tests in audiometric booths. *Noise Control*, 7, 22-26 (1961)

[3] Bass, H. E. , Sutherland, L. C. , Piercy, J., and Evans, L. (1984). Absorption of sound by the atmosphere. *Physical Acoustics*, 17, 145–232.

[4] Bentz, C. and Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, 3:1–27.

[5] Boncoraglio, G. and Saino, N. (2007). Habitat structure and the evolution of bird song. *Functional Ecology*, 21(1):134–142.

[6] Brown, A. M. and Lindsey, D. T. (2004). Color and language: Worldwide distribution of Daltonism and distinct words for blue. *Visual neuroscience*, 21(03):409–412.

[7] Chater, N. and Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, 34(7):1131–1157.

[8] Coupé, C. (2015). Testing the acoustic adaptation hypothesis with GIS and spatial regression models. *Causality in the Language Sciences Workshop*, Leipzig.

[9] Croft, W. (2000). *Explaining language change: an evolutionary approach*. Harlow, Essex: Longman.

[10] Dediu, D. and Ladd, D. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences*,104(26):10944.

[11] Donohue M, Hetherington R, McElvenny J, Dawson V (2013) *World Phonotactics Database*. Available at http://phonotactics.anu.edu.au/.

[12] Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

[13] Dunbar, R. (1993). Coevolution of neocortical size, group-size and language in humans. *Behavioral and Brain Sciences*, 16(4):681–694.

[14] Dunn, M., Greenhill, S. J., Levinson, S. C., and Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473(7345):79–82.

[15] Ember, C. and Ember, M. (2007). Climate, econiche, and sexuality: Influences on sonority in language. *American Anthropologist*, 109(1):180–185.

[16] Engwall, O., Delvaux, V., and Metens, T. (2006). Interspeaker variation in the artic- ulation of nasal vowels. *Proceedings of the 7th ISSP*, p. 3–10.

[17] Everett, C., Blasi, D. E., and Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences*, 122(5): 1322–1327

[18] Filippi, P., Gingras, B., and Fitch, W. T. (2014). Pitch enhancement facilitates word learning across visual contexts. *Frontiers in psychology*, 5.

[19] Fought, J. G., Munroe, R. L., Fought, C. R., and Good, E. M. (2004). Sonority and climate in a world sample of languages: Findings and prospects. *Cross-cultural research*, 38(1):27–51.

[20] Honkola, T., Vesakoski, O., Korhonen, K., Lehtinen, J., Syrjänen,K., & Wahlberg,N. (2013). Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology*, 26(6):1244–1253.

[21] Hunter, M. L. and Krebs, J. R. (1979). Geographical variation in the song of the great tit (parus major) in relation to ecological factors. *The Journal of Animal Ecology*, pages 759–785.

[22] Isbell, L. A. and Young, T. (1996). The evolution of bipedalism in hominids and reduced group size in chimpanzees: alternative responses to decreasing resource availability. *Journal of Human Evolution*, 30:389–397(9).

[23] Koptjevskaja-Tamm, M. and Rakhilina, E. V. (2006). Some like it hot: On the semantics of temperature adjectives in russian and swedish. *Sprachtypologie und Universalienforschung*, 59(3):253–269.

[24] Lupyan, G. and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, 5(1):e8559.

[25] Moisik, S. R., Lin, H., & Esling, J. H. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association*, 44, 21-58.

[26] Moisik, S., & Gick, B. (2013). The quantal larynx revisited. *The Journal of the Acoustical Society of America*, 133(5), 3522-3522.

[27] Morgan, L . C . Effect of temperature and humidity on hearing acuity. *Volta Review*, 62, 364-68 (1960)

[28] Morton, E. S. (1975). Ecological sources of selection on avian sounds. *American Naturalist*, pages 17–34.

[29] Nettle, D. (1999). *Linguistic Diversity*. Oxford: Oxford University Press.

[30] Noback, M. L., Harvati, K., and Spoor, F. (2011). Climate-related variation of the human nasal cavity. *American journal of physical anthropology*, 145(4):599–614.

[31] Oghan F, Cingi C, Seren E, Ural A, Guvey A. (2010) Assessment of the impact of altitude on nasal airflow via expiratory nasal sound spectral analysis. *European Archives of Oto-Rhino-Laryngology*., 267(11):1713-8.

[32] Richerson, P. J., Boyd, R., and Bettinger, R. L. (2009). Cultural innovations and demographic change. *Human Biology*, 81(3):211–235.

[33] Schuit, J. (2012) Signing in the Arcic: External influences on Inuit sign language. In Zeshan, U. and de Vos, C., editors. *Sign languages in village communities*: Berlin: Mouton de Gruyter.

[34] Snell-Rood, E. C. (2012). The effect of climate on acoustic signals: does atmospheric sound absorption matter for bird song and bat echolocation? *The Journal of the Acoustical Society of America*, 131(2):1650–1658.

[35] Steels, L. and Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: a case study for colour. *Behavioral and Brain Sciences*, 28(4):469–489.

[36] Tallavaara, M. & Seppä, H. (2011). Did the mid-Holocene environmental changes cause the boom and bust of hunter-gatherer population size in eastern Fennoscandia? *The Holocene*, 22(2):215–225.

[37] Tamariz, M., Cornish, H., Roberts, S., and Kirby, S. (2012). The effect of generation turnover and interlocutor negotiation on linguistic structure. In T. C. Scott-Phillips et al., editor, *The Evolution of Language: Proceedings of the 9th International Conference*, pages 555–556. World Scientific.

[38] Torreira, F., Roberts, S. G., and Hammarström, H. (2014). Functional trade-off between lexical tone and intonation: typological evidence from polar-question marking. In *The 4th International Symposium on Tonal Aspects of Language*.

[39] Warren, P. S., Katti, M., Ermann, M., and Brazel, A. (2006). Urban bioacoustics: it's not just noise. *Animal behaviour*, 71(3):491–502.

[40] MacWhinney, B. (2005). Language evolution and human development. In D. Bjorklund & A. Pellegrini (Eds.), *Child development and evolutionary psychology*. New York: Academic.

GENERAL PURPOSE COGNITIVE PROCESSING CONSTRAINTS AND PHONOTACTIC PROPERTIES OF THE VOCABULARY

Padraic Monaghan[1] & Willem H. Zuidema[2]
[1]Department of Psychology, Lancaster University, Lancaster UK
[2]Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam The Netherlands

ABSTRACT

Properties of phonological systems may derive from both comprehension and production constraints. In this study, we test the extent to which general purpose constraints from sequence production are manifested in repetitions of phonemes within words. We find that near repetitions of phonemes occur less than expected by chance within the vocabularies of four studied languages: Dutch, English, French and German. This is consistent with constraints on response suppression effects in short term sequence production and with the principle of "similar place avoidance", but inconsistent with theories of consonant harmony derived from formalisation of co-articulation of phonemes in speech.

Keywords: phonotactics, production constraints, repetition, consonant harmony, co-articulation.

1. INTRODUCTION

An influential and effective approach to studies of evolutionary adaptation in structuring phonology has been to explore comprehension pressures for signal detection. For instance, Zuidema and de Boer [15] demonstrated that combinatorial phonology was an effective solution to maximizing fidelity in the acoustic form of words, and Kirby et al. [8] showed in studies of language transmission that words were likely to change from holistic forms to incorporate combinatorial structure to maximize comprehension efficiency. Similarly, Monaghan et al. [11] showed that the arrangement of form-meaning mappings in language was more easily acquired if the relation between sound and meaning was arbitrary rather than systematic. In context, an arbitrary mapping enabled the maximum information from the signal to be used in order to disambiguate the intended referent, thus enhancing the distinctiveness and signal to noise ratio for comprehension.

Each of these approaches has shown that general purpose learning mechanisms applying to comprehension result in observed patterns of phonological structure. But other properties of phonology may be the result of an accumulation of pressures from general purpose constraints on production. In this paper we address one such potential production limitation: the occurrence of repeating phonemes in the vocabulary. The distribution of repetitions within words provides insight into the communicative pressures that have resulted in the phonotactic patterns observed in extant vocabularies.

Phonological productions require a sequence of phonemes to be articulated, and as such they are prone to general purpose production constraints on sequences. One possible influential constraint is the effect of repetition on sequence encoding and/or reproduction from the memory literature. In short term memory tasks, if participants are required to recall a sequence containing a repetition then the consequence is a reduction in recall accuracy for the repeated number, particularly when it is separated by 1, 2, or 3 other numbers. This observation, known as the Ranschburg effect [3, 5], has been linked to constraints on production, as the reproduction of the sequence during recall is prone to response suppression which prohibits the same element being reproduced more than once [7]. This process is likely to result in *fewer* repetitions within words of phonemes than expected by chance.

A potentially counteractive pressure from production results from co-articulation effects, that assimilates manner or place of articulation of phonemes at points close together in the speech signal [4, 14]. This general purpose constraint on production would have the consequence that repetitions of phonemes may occur *more* than expected by chance.

What is currently lacking in the literature is a comprehensive quantitative analysis of phoneme repetitions at different positions within words in the vocabulary in order to determine whether either of these potential production constraints are affecting the phonotactic structure of the vocabulary. Given that phoneme inventories, and phonotactic constraints, co-evolve to address the joint issue of maximising perception but minimizing production effort. Thus, investigating how such constraints may relate to general purpose cognitive or articulatory limitations is key to understanding extant phonemic inventories, syllabic structure, as well as the way such phonotactic constraints can be used to support word identification, e.g., [2].

One exception is a previous study of repetition distributions in the work of MacKay [9], who showed that for subsamples of Croatian and

Hawaiian, there appeared to be a peak of repetitions for vowels one phoneme apart, and a peak for repetitions of consonants 3 phonemes apart. However, these studies were on only small samples of the corpora, and the extent to which other phonotactic constraints were driving the effects – such as the sonority hierarchy – were not possible to discern in these small-scale analyses.

Related to this is a cross-linguistic analysis of "similar place avoidance", where pairs of consonants within words are less likely to have the same place of articulation [13]. Across 30 languages, there are fewer attested forms of words containing the same place of articulation for pairs of consonants. However, the distance between phonemes that contributed to the similar place avoidance was not determined, and nor was its relation to other properties of phonemes, such as similarities in manner of articulation.

We address the issue of whether repetitions are more or less likely than chance, where we take various other constraints into account in determining a baseline, random distribution of repetitions. If the co-articulation harmony hypothesis affects phonotactic structure then we would anticipate a greater number of repetitions between phonemes in the vocabulary than chance, whereas if the Ranschburg effect affects phonotactic structure of the vocabulary, then we would expect that repetitions of phonemes close together in the word occur at a frequency less than chance.

Table 1: Properties of the vocabularies used in the analyses.

Property	Dutch	English	French	German
Number of words	117,116	53,699	62,123	79,675
Mean word length (phonemes)	9.090	6.970	6.852	8.890
Number of distinct phonemes	44	53	39	57

2. CORPUS PREPARATION

We investigated four different languages: Dutch, English, French, and German. The vocabulary lists were taken from the CELEX database [1] for Dutch, English, and German, and from Lexique 3.80 [12] for French. Only forms that were attested in the corpora used to generate frequency information were included (so for the CELEX lists, and for Lexique where frequency information was taken from the film under-titles database, frequency was > 0). Lists of words included only unique phonological forms, so homophones occurred only once in each vocabulary. Forms that comprised more than one word in the databases were also omitted, but compound forms that were listed as a single word were included. Table 1 shows the characteristics of each vocabulary.

3. REPETITIONS OF CONSONANTS AND VOWELS

3.1. Analysis

For each vocabulary, we investigated the within-word repetitions at different distances, from 1 (adjacent phonemes) to 10 (with 9 intervening phonemes) phonemes apart. Note that for larger distances, only the longer words in the vocabulary would be contributing to the counts. At each distance, the number of repetitions of phonemes was assessed. These were separated into repetitions containing consonants and those containing pairs of vowels, in order to account for different phonotactic constraints applying to vowels and consonants – i.e., vowels tend to be preceded and followed by consonants [6]. Thus, for the word "popular" (/pɒpjʊlə/), at separation distance of one, the consonant pair /pj/ would be assessed for repetitions, and at this distance this resulted in no vowel-vowel pairs. For the vowels, at this separation distance, the word contributed no repetitions. Then, for distance of 2, the pairs /pp/ and /jl/ for the consonants, and /ʊə/ for the vowels would be assessed. At this separation distance, the word contributed one repetition in the consonant analysis (/pp/). Then, repetitions at separation distance of 3 were calculated, and so on up to distance of 10 phonemes (though for the word "popular", there were no phoneme pairs assessed beyond separation distance of 6 phonemes).

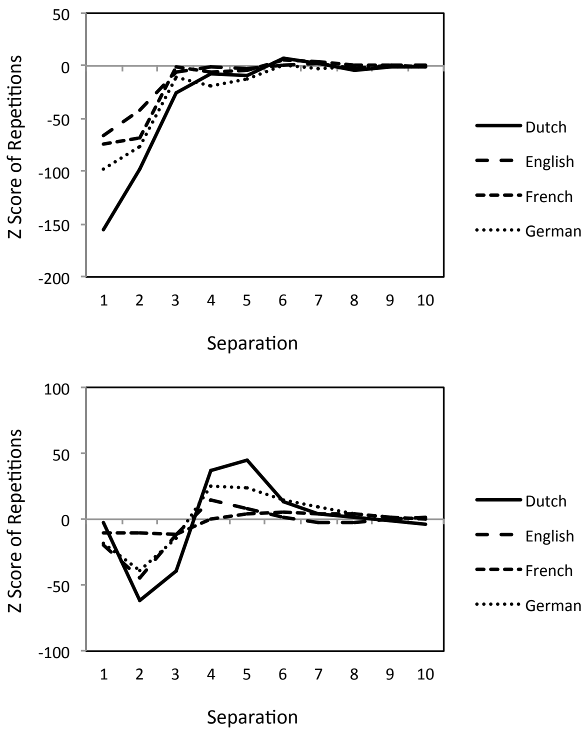
In order to determine whether the repetitions in the vocabulary occurred with frequency greater or less than expected by chance, we compared the actual number of repetitions to a random baseline distribution, where the phonemes within consonant pairs or within vowel pairs at each separation distance were randomly reassigned and the number of repetitions that occurred by chance was determined. This was repeated 10,000 times in a Monte Carlo analysis.

The resulting baseline distributions were similar to normal distributions, and so we determined the Z-score of the actual repetitions that occurred against the random distribution. Z-scores less than 0 indicate actual repetitions occur less than expected by chance, Z-scores greater than 0 indicate actual repetitions are greater than expected by chance. Z-scores > |2.81| are significantly different than chance ($p < .05$).

3.2. Results

Figure 1 shows the results of the analysis of consonant repetitions and vowel repetitions, respectively, within each vocabulary. The x-axis indicates the number of phonemes intervening between the repetition. The y-axis indicates the Z-score of the actual number of repetitions against the repetitions resulting from randomised versions of the corpus.

Figure 1: Repetitions of consonants (upper) and vowels (lower).



The results were very similar across all the languages. Not unexpectedly, there were fewer immediate repetitions than expected by chance for both vowels and consonants (separation distance 1). However, this suppression of repetition also pertained for separations up to 5 apart for the consonants, and 3 apart for vowels. For consonants separated by more than 5 other phonemes, there was variation across the languages for whether repetitions were at chance, or slightly above chance. Both Dutch and French had more repetitions than expected by chance at separation distances 5 and 6, and English and German were not significantly different than chance.

For the vowels, there was a general pattern of greater repetitions than expected by chance for separation distances 4 to 6. For longer separation distances, the distribution of repetitions converged to chance levels.

The general pattern of repetitions observed in these four languages is somewhat consistent with that of MacKay's [9] analyses of small subsets of corpora in Croatian and Hawaiian, and is in alignment with general cognitive processing constraints that drive the Ranschburg effect in short term memory tasks. Thus, across these languages, there are fewer instances of words such as "bob" or "blob", than there are words without repetitions such as "bod" or "blot".

Table 2: Mock example of calculating, separated by one other phoneme, consonant repetitions, repetitions modulated by manner, and repetitions modulated by place of articulation. Note for randomised same manner, phonemes are randomised across sets with the same manner of articulation (so only phonemes in the pairs p_g, t__p, b_b, and b_p are interchangeable, and v_v is only interchangeable with itself). For randomised same place, only phonemes in the pairs p_b and b_p are interchangeable, s_t is only interchangeable with itself, and v_v is only interchangeable with itself.

Word	Con pairs	Ran Con pairs	Same mann. pairs	Ran Same mann.	Same place pairs	Ran Same place
pop	**p_p**	**p_p**	**p_p**	p_g	**p_p**	p_b
sot	s_t	s_b			s_t	s_t
top	t_p	t_v	t_p	t_p		
bob	**b_b**	b_t	**b_b**	**b_b**	**b_b**	**b_p**
bog	b_g	b_p	b_g	b_p		
viv	**v_v**	v_g	**v_v**	**v_v**	**v_v**	**v_v**
Total Reps	3	1	3	2	3	2

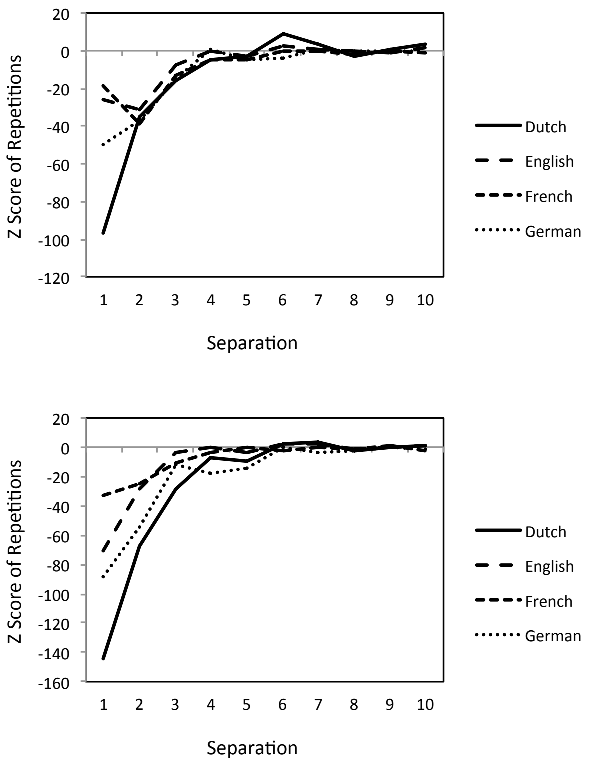
4. REPETITIONS OF CONSONANTS MODULATED BY MANNER AND PLACE

4.1. Analysis

The previous analyses of consonants take as a random baseline any two consonants, and reorder their co-occurrences. However, there are phonotactic constraints that operate over phonemes in terms of their manner of articulation that are due to syllabic structure rather than other limitations on phoneme occurrences within these structures. For instance, the sonority hierarchy permits plosive-approximant sequences in onsets but not in codas of syllables. To better respect these potential constraints that derive from the syllable structure, we repeated the analyses of consonant repetitions, but differentiated repetitions according to manner of articulation. Thus, only the phoneme pairs with similar manner of articulation were considered and random reassignments of the phonemes to these pairs occurred within manner of articulation pairs. So, for the example /pɒpjʊlə/, at separation distance 2 for the plosive manner of articulation only /p-p/ contributed to the set of plosive pairs to be reassigned, and only /j-l/ contributed to the set of approximant pairs to be reassigned. These randomized sublists were then tested for repetitions and the results were summed and compared to the actual repetitions occurring in the vocabulary. Table 2 shows a mock example of the calculations.

A similar analysis was performed but this time considering pairs of consonants that had the same place of articulation (so for the /pɒpjʊlə/ example, /p-p/ would be entered into the set of bilabial phonemes for random reassignment but /j-l/ would not be included in a randomized set because the place of articulation differed (see Table 2 for an example). The effect of these analyses modulated by manner or by place was to inflate the repetitions of phonemes that occurred by chance in the Monte Carlo randomized analyses.

Figure 2: Repetitions of consonants modulated by manner (upper) and place (lower) of articulation.



4.2. Results

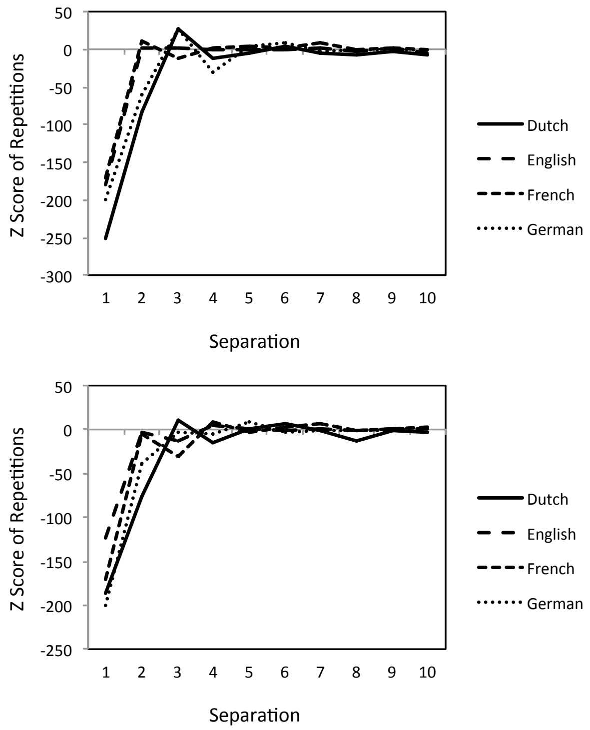
The results are shown in Figure 2 for consonant repetitions modulated by manner of articulation, and modulated by place of articulation. The effects generally reflect the previous analyses: repetitions occur less than expected by chance for consonants that occur close together in words. The avoidance of place similarity for near consonants reflected that of previous studies of the similar place avoidance principle [2, 12] but showed in addition that separation distance weakened the avoidance effect.

5. REPETITIONS OF ABSTRACT STRUCTURE

5.1. Analysis

The analyses thus far have assumed that constraints on repetition apply at the phoneme level. Thus, it is still possible for consonant harmony effects to be observed, which apply more abstractly to classes of phonemes with similar place or manner of articulation. Hence, repetitions of individual phonemes could be inhibited, but still repetitions of phonemes of the same manner of articulation could occur more than expected by chance. This would be a way in which consonant harmony effects could co-exist with reduced repetition of individual phonemes. In the final set of analyses, we assessed the extent to which repetitions of phonemes with the same manner of articulation were repeated at different separations within the vocabulary. Thus, if a plosive occurred with any other plosive that would be counted as an occurrence of a repetition. The random baseline was computed by randomly assigning phonemes to positions, but then measuring the manner of articulation of these randomly rearranged vocabularies. A similar analysis was conducted for phonemes with the same place of articulation.

Figure 3: Repetitions of phonemes with same manner (upper) and place (lower) of articulation.



5.2. Results

Figure 3 shows the results for repetitions of phonemes with the same manner of articulation and for phoneme repetitions with the same place of articulation. The general pattern of results demonstrate that phonemes with the same manner or place of articulation tend to be inhibited at near positions in the vocabulary. However, there are some exceptions. For English, there are slightly more repetitions of phonemes with the same manner of articulation with one other phoneme separating (so "bod" is more likely than "mod"). For Dutch and German, there is a peak at distance 3 for the same manner of articulation (so "dank" would be more likely than "rank"). For place of articulation, the only repetition that occurs more than chance is for Dutch at separation distance 3. Thus, there may be some small contributions of consonant harmony effects for some of these languages, but the general effect is that there are reduced co-occurrences of phonemes with the same manner or place of articulation, again consistent with the similar place avoidance principle [12], but again that it is a graded phenomenon according to distance.

6. CONCLUSION

The starting point for these analyses was to determine whether repetitions occurred more or less than by chance, to test whether phonotactic structure was consistent with either the immediate suppression of repetitions as predicted by the Ranschburg effect, or enhancement of repetitions as predicted by co-articulation accounts of consonant harmony. The general results are more consistent with the former general purpose production constraint: close repetitions of phonemes are less likely than expected by chance within the vocabularies of the four languages we have analysed. However, this suppression effect appeared to (also) operate more abstractly in terms of suppressing repetitions of phonemes with the same manner or same place of articulation, consistent with the similar place avoidance principle [2]. Thus, there are in fact fewer consonant harmony effects than expected by chance at close distances of separation. One possible explanation for this is that co-articulation effects are actually inhibited in the vocabulary to prevent mistaken apprehension of co-articulatory effects: If the speaker produces a co-articulation then the listener can be sure that this is an error of production, therefore avoiding possible ambiguities of production [10].

These corpus analyses provide a first step to establishing the phenomena within the phonotactic structure of these languages. The next step is to confirm with experimental studies the effect of repetitions of phonemes and classes of phonemes at near and far points of repetition in the vocabulary.

7. REFERENCES

[1] Baayen, R.H., Pipenbrock, R. & Gulikers, L. (1995). *The CELEX Lexical Database* (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

[2] Boll-Avetisyan, N.A.T. & Kager, R.W.J. (2014). OCP-PLACE in speech segmentation. *Language and Speech*, *57*, 394-421.

[3] Crowder, R. G. (1968). Intraserial repetition effects in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, *7*(2), 446-451.

[4] Hansson, G. (2001). The phonologization of production constraints: Evidence from consonant harmony. *Chicago Linguistic Society* (Vol. 37, p. 200).

[5] Henson, R. N. (1998). Item repetition in short-term memory: Ranschburg repeated. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1162.

[6] Hockema, S. A. (2006). Finding words in speech: An investigation of American English. *Language Learning and Development, 2*, 119-146.

[7] Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, *27*(2), 117-143.

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

[9] MacKay, D. W. (1970). Phoneme repetition in the structure of languages. *Language and Speech*, *13*, 199-213.

[10] McRoberts, G. W., McDonough, C., & Lakusta, L. (2009). The role of verbal repetition in the development of infant speech preferences from 4 to 14 months of age. *Infancy*, *14*(2), 162-194.

[11] Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The Arbitrariness of the Sign: Learning Advantages From the Structure of the Vocabulary. *Journal of Experimental Psychology*, 325-347.

[12] New B., Pallier C., Ferrand L., Matos R. (2001) Une base de données lexicales du français contemporain sur internet: LEXIQUE, L'Année Psychologique, 101, 447-462. http://www.lexique.org

[13] Pozdniakov, K., & Segerer, G. (2007). Similar Place Avoidance: A statistical universal. *Linguistic Typology*, *11,* 307–348.

[14] Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, *80*, 475-531

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

SIMULATING THE INTERACTION OF FUNCTIONAL PRESSURES, REDUNDANCY AND CATEGORY VARIATION IN PHONETIC SYSTEMS

Bodo Winter[1] & Andy Wedel[2]

[1] University of California, Merced, Cognitive and Information Sciences
[2] University of Arizona, Department of Linguistics

ABSTRACT

Phonetic systems need to be able to signal communicatively relevant meaning distinctions. In this paper, we explore an evolutionary simulation which shows how the functional pressure to keep words perceptually distinct reduces variation at the phonetic level. Our simulation furthermore shows that adding redundancy to the system (e.g., through multiple phonetic cues or longer words) relaxes these functional pressures. Based on these results we argue that phonetic systems can be seen as finding a relative optimum: Efficient and unambiguous communication is maintained while at the same time, there is enough category variation to allow evolvability, the potential for future evolution.

Keywords: redundancy; category variation; evolutionary simulations; exemplar models; cultural evolution

1. INTRODUCTION

The sound systems of spoken languages constantly change, at both long and short time scales [1, 27, 8]. What remains constant amidst these changes is the ability of sound systems to subserve communication [39]. A broad range of work argues that this arises in part through functional pressures on language evolution to maintain sufficient contrast in phonetic categories [3, 13, 21, 29, 33, 34, 41]. In this paper, we look at how phonological systems evolve under such usage constraints.

We specifically investigate the role of a functional pressure towards keeping words acoustically distinguishably (henceforth "anti-ambiguity bias"). We suggest that this bias constrains variation at the phonetic level, i.e., different renderings of the same utterance vary less if lexical items need to be contrasted. Moreover, our simulations show that infusing redundancy into phonetic systems (e.g., via multiple phonetic cues or via longer words) relaxes this functional pressure, allowing systems to harbour more variation.

2. BACKGROUND

Phoneme contrasts such as /p~b/, /s~f/, or /ɑ~ɔ/ can be lost from a language when, for example, two phonemes merge with one another [16, Ch. 11]. As an example, the contrast between /ɑ~ɔ/, (exemplified by the words "cot" and "caught") has merged in many dialects of North America [17]. Wedel, Jackson, and Kaplan [35] demonstrated that the probability of such merger is cross-linguistically associated with how many lexical items are distinguished by the phonemic contrast: a greater number of such "minimal pairs" is significantly associated with lower merger probability (see also [36]).

The linguistic literature is rife with anecdotal reports of these kinds of effects as well. Blevins and Wedel [2] discuss attested cases of "inhibited" sound changes, where an otherwise regular sound change ignores sets of words that would lead to the breakdown of an entire morphological paradigm, such as the distinction between past tense and present tense. These kinds of observations suggests that biases toward communicative efficiency do influence the evolution of phonological systems. Moreover, they corroborate the statistical studies [35,36] which indicate that maintenance of contrast at the lexical level in particular, influences the phonetic level.

Wedel [34] proposes a multi-level exemplar model to account for these interactions (cf. [31]). In exemplar models of speech, phonological knowledge is characterized as being constituted by rich and detailed representations of experience rather than by abstract symbolic representations. Within this framework, a phonological category can be modeled as a collection of stored phonetic exemplars (an "exemplar cloud") that is acquired and continually enriched by experience.

Language evolution can then be modelled as resulting from a repeated cycle of production and perception events [26, 33]. To model interactions between the lexicon and sublexical structure, experiences need to contribute to two connected levels of representation: a lexical level, and a sublexical level (phonetics). In Wedel's computational model [32, 33, 34], an anti-ambiguity bias at the lexical level results in the evolution of a phoneme set that efficiently subserves lexical distinctions.

This paper extends this model to explore what has been called "cryptic variation" in biological systems [10, 30]. Cryptic variation refers to variation that is not selected for or against, that is, neutral variation that does not impact fitness. For language, it has been noted that sound systems harbour variation that is not consciously perceived by speakers [25], and therefore is not subject to overt communicative pressures.

We can use the concept of cryptic variation to understand an observation that has frequently been made by linguists: Sound systems that do not make use of certain distinctions tend to "allow for" or "afford" more variation. For example, Lavoie [19] shows that native speakers of English produce spirantized variants of /k/ more frequently than native speakers of Spanish, where /k/ and /x/ are contrastive. Thus, the English sound category of /k/ encroaches into "unfilled areas of the language's sound space" [19, p. 39]. However, in Spanish variation is more constrained, presumably because of the functional significance of /x/ in that particular language's system. This reduction of variation due to communicative significance is what we set out to model.

3. THE COMPUTATIONAL MODEL

This section briefly outlines the computational model, with more technical detail provided in [34]. In the model, two agents take turns talking to each other. Each agent has an internal lexicon. The speaker utters one token of each of the words, and the listener maps each token to its best fitting category, where it stores the input as a new exemplar.

Each word exemplar is further decomposed into a number of phonetic exemplars on one of two possible continuous dimensions, each with an arbitrary scale from 1-100. As a useful metaphor, we can think of one of the dimensions as voice-onset time (VOT), and the other dimension as vowel height on an /i-a/ continuum. Thus, each word exemplar maps onto a point in 2-dimensional space. For example, a token with the values [15 VOT, 25 TongueHeight] can be thought of as corresponding to [ba].

Each new exemplar is associated with an initial activation value that decreases over time, corresponding to the observation that memories decay [11, 15, 24, 26]. In production, exemplars are selected as a function of the activation level, with more strongly activated exemplars contributing more strongly to a production plan.

For each word production, a random exemplar is chosen from the word's exemplar cloud. Two types of changes apply to the target before it is passed to the listener for categorization: the addition of production noise, and the application of a similarity bias [26], reviewed in [33]. This similarity bias is implemented by biasing the phonetic values of the output target toward nearby values in memory both within the word category itself, and across the lexicon. The move towards nearby values of the lexicon leads the system to re-use phonetic features across words, which is a defining characteristic of human languages [12, 18, 20]. There is empirical support for such a cross-word similarity bias, reviewed in [34].

A final feature of the model is a bias against lexical confusability [34]. A bias with this effect is empirically motivated by the above-mentioned cross-linguistic studies of phoneme merger and inhibited sound change. We implement this bias computationally in a straightforward fashion: an output has a chance of not being stored as a new exemplar in the listener's memory in proportion to the degree to which it maps to multiple categories [32, 33]. In this way, unambiguous speaker outputs are more likely to be stored than ambiguous outputs, with the result that unambiguous exemplars contribute relatively more to the continuing evolution of the lexicon. The central result from this work is that contrast relationships between sounds may be constrained and maintained by contrast relationships between words.

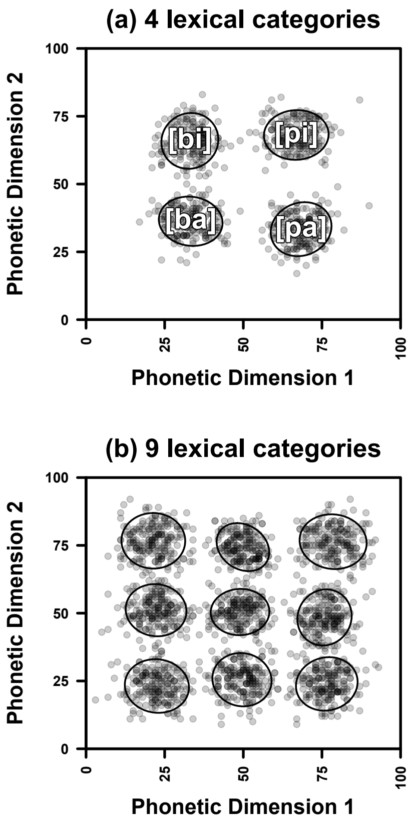
4. SIMULATION RESULTS

4.1. Category variation as a function of lexical density

We first explored how lexical density impacts cryptic variation at the phonetic level. Figure 1 shows a time slice of two representative simulations after 500 time steps, where agents either had 4 words or 9 words. Note that these are extremely small lexicons compared to natural languages because we are specifically interested in modelling elements of the lexicon that form a set of minimal pairs.

In the absence of other words, a lexeme's exemplar distribution is determined by the balance between noise, which promotes spread, and similarity bias, which promotes contraction [26]. When the exemplar clouds of two words get close enough such that some outputs become ambiguous in perception, the anti-ambiguity bias comes into play as well, which introduces an additional constraint on how broad a category can spread. This anti-ambiguity bias is stronger when there is higher lexical density. In other words, as we add more words in a given phonetic space, pronunciation variation at the boundaries between them becomes increasingly suppressed and the standard deviations of the exemplar clouds shrink. This can be seen in Figure 1, where the dashed lines indicate the standard deviations (SD) of all exemplar clouds for different numbers of lexical categories (after 500 simulation steps).

Figure 1: Simulation results with (a) 4 words and (b) 9 words, after 500 time steps. Each point represents an exemplar. Each cloud represents the totality of exemplars for a word. Labels are given for ease of interpretation. Ellipsoids represent confidence regions that cover 80% of the exemplars for each cloud.

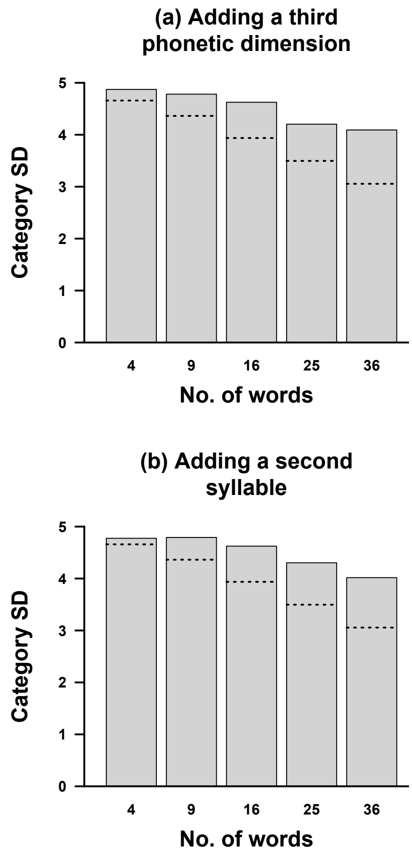


4.2. Increasing redundancy

It is well known that phonological categories are often distinguished by many different phonetic cues [9, 14, 28, 40], which has been argued to increase the robustness of speech communication [39]. What if we add another dimension, akin to having another phonetic cue?

We introduced an additional independent dimension. For ease of interpretation, we can imagine this to be an additional vowel contrast along the front-back dimension allowing an expanded vowel space of /i ~ ɯ ~ æ ~ ɑ/. The information provided by the additional phonetic cue allows for the maintenance of variation in the initial dimension of each lexical item even as more lexical items are added. In Fig 2a, each bar represents the SD with an added third dimension of phonetic contrast, while dashed lines indicate the SD of corresponding simulations without this third dimension. Thus, adding a third dimension increases standard deviations. This is because with this additional phonetic dimension, the role of each phonetic cue in reducing lexical confusability is reduced, which relaxes constraints on spread.

Figure 2: The relationship between category standard deviation and lexical density for systems with added redundancy through (a) a third phonetic dimension or (b) a second CV syllable. Dashed lines indicate SDs of simulation runs without such redundancy.



4.3. Making words longer

In this set of simulations, we limit the phonetic space to the two dimensions used in section 4.1, but double the length of each word by copying the values of the first two dimensions to a second "syllable" when initializing the simulation. After the start of the simulation, the numeric values of this second syllable are independent of those in the first syllable, except insofar as they belong to the same dimension and so are subject to the same phonetic similarity bias. Hence, the second syllable potentially adds just as much information about the output identity as the first. By increasing the amount of phonetic material transmitted (in terms of word length), but keeping the number of lexemes the same, we by definition increase redundancy. Again, similar to the case of adding a third dimension, the constraint on category standard deviation is relaxed (see Figure 2b) and SDs are higher.

5. DISCUSSION

The present simulations illustrate that lexical density directly affects variation within this model architecture, as expected. Exemplar clouds become more constrained when adding more possible lexical items. Within this model, the mechanism by which this happens is an anti-ambiguity bias, the same bias that has been proposed to explain patterns of phoneme merger [35, 36].

This anti-ambiguity bias directly relates to a listener's uncertainty about the incoming input. This is clearly demonstrated by adding redundancy to the system, which increases global phonetic distance between signals, rendering them less confusable. This is a concrete example of how redundancy serves to counteract noise [7].

Crucially, once redundancy has been added to the system —either via additional phonetic cues or via longer words— exemplar clouds are less constrained in their cryptic variation, and within-category variation is allowed to accumulate. This within-category variation is crucial for future change, as all evolution needs variation as "fodder". Hence, increasing redundancy increases variation and hence, assures the future evolvability of the system. Wedel [32, section 3.3] illustrates how variation provides a pathway for sound change in this architecture.

A comment should be made about the term "redundancy". From an information-theoretical perspective, adding a new syllable or adding an additional phonetic dimension are qualitatively similar changes; they both expand the phonetic channel capacity through which lexical contrasts can be distinguished, beyond what is strictly speaking necessary to distinguish lexemes. In the disyllabic case, redundancy is added in a sequential fashion. In the phonetic dimensionality case, redundancy is added in a simultaneous fashion. In the biological literature, each of these types of redundancy is technically called "degeneracy" (for review, see [23]), which refers to redundancy in which *different* structural components realize similar system functions [4, 22, 23, 37, 38]. In the linguistic case, this corresponds to different syllables and different phonetic cues signalling the same contrast. Even if the same syllable is repeated, this does strictly speaking not fall under the purview of redundancy, because the syllable conveys linguistic information at a different time point.

The present results are conceptually important because they show how evolution at one level (the lexicon) affects evolution at another level (the phoneme system). This deviates from standard exemplar models in the domain of speech, e.g. Goldinger [5], who models words as holistic acoustic traces—there are no separate sublexical and lexical levels in his model. However, a two-layer exemplar architecture is necessary in the present case to model lexicon/speech interactions.

This two-level architecture is furthermore illustrative because there are direct parallels to biological evolution, where evolution acts on phenotypes, and therefore selection only indirectly affects the frequency of genotypes within a population. Similarly, in the linguistic case modelled in the present paper, selection acts indirectly on phoneme inventories, via coupling relations from the lexical to the sublexical levels. This general picture is moreover in line with the idea that the "success" of phonetic categories is largely measured with respect to what they do at the communicative level [6, 35, 36]. The communicative "currency" in this model, so to say, is the word, not the phoneme. And this currency is ultimately the measure of success for different phonetic exemplars.

7. REFERENCES

[1] Blevins, J. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge, UK: Cambridge University Press.

[2] Blevins, J., Wedel, A. 2009. Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26, 143–183.

[3] de Boer, B. 2000. Self-organization in vowel systems. *Journal of Phonetics* 28, 441–465.

[4] Edelman, G. M., Joseph A. G. 2001. Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences* 98, 13763–13768.

[5] Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological review* 105, 251–279.

[6] Graff, P. 2012. Communicative efficiency in the lexicon. Doctoral dissertation, Massachusetts Institute of Technology.

[7] Hailman, J. P. 2008. *Coding and redundancy: Man-made and animal-evolved signals*. Cambridge: Harvard University Press.

[8] Harrington, J., Palethorpe, S., Watson. C. I. 2000. Does the Queen speak the Queen's English? *Nature* 408, 927–928.

[9] Hawkins, S. 2010. Phonological features, auditory objects, and illusions. *Journal of Phonetics* 38, 60–89.
[10] Hayden, E J., Ferrada, E., Wagner, A. 2011. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474, 92–95.
[11] Hintzman, D. L. 1986. 'Schema abstraction' in a multiple-trace memory model. *Psychological Review* 93, 328–338.
[12] Hockett, C. 1960. The origin of speech. *Scientific American* 203, 88–96
[13] Hockett, C. 1967. The quantification of functional load. *Word* 23, 320–339.
[14] Kingston, J., Diehl, R. D. 1994. Phonetic knowledge. *Language* 70, 419–454.
[15] Kruschke, J. K. 1992. ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* 99, 22–44.
[16] Labov, W. 1994. *Principles of linguistic change: Internal factors*. Oxford: Wiley-Blackwell.
[17] Labov, W., Ash, S., Boberg, C. 2006. *The Atlas of North American English: Phonetics, phonology, and sound change*. Berlin: Mouton de Gruyter.
[18] Ladd, R. 2013. What is duality of patterning, anyway? *Language and Cognition* 4, 261–273.
[19] Lavoie, L. M. 2002. Subphonemic and suballophonic consonant variation: The role of the phoneme inventory. *ZAS Papers in Linguistics* 28, 39–54.
[20] Lindblom, B., Diehl, R. K., Park, S.-H., Salvi, G. 2011. Sound systems are shaped by their users: The recombination of phonetic substance. In: Nick, G. C., Ridouane, R. (eds.), *Where do phonological features come from?: Cognitive, physical and developmental bases of distinctive speech categories*. Amsterdam: John Benjamins, 67–98.
[21] Martinet, A. 1952. Function, structure, and sound change. *Word* 8, 1–32.
[22] Mason, P. 2010. Degeneracy at multiple levels of complexity. *Biological Theory* 5, 277–288.
[23] Mason, P, Domínguez D., J. F., Winter, B., Grignolio, A. 2015. Hidden in plain view: Degeneracy in complex systems. *BioSystems* 128, 1–8.
[24] Nosofsky, R. M. 1986. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115, 39–57.
[25] Ohala, J. 1989. Sound change is drawn from a pool of synchronic variation. In: Breivik, Leiv E., Jahr, Ernst H. (eds.), *Language Change: Contributions to the study of its causes*. Berlin: Mouton de Gruyter.
[26] Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee, J. L., Hopper, P. J. (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 137–158.
[27] Sancier, M. L., Fowler, C. A. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25, 421–436.
[28] Stevens, K. N., Keyser, S. J. 2010. Quantal theory, enhancement and overlap. *Journal of Phonetics* 38, 10–19.
[29] Trubetzkoy, N. 1939. *Grundzüge der Phonologie*. Prague, Czech Republic: Travaux du Cercle Linguistique de Prague.
[30] Wagner, A. 2005. *Robustness and evolvability in living systems*. Princeton: Princeton University Press.
[31] Walsh, M., Möbius, B., Wade, T., Schütze, H. 2010. Multilevel exemplar theory. *Cognitive Science* 34, 537–582.
[32] Wedel, A. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23, 247–274.
[33] Wedel, A. 2007. Feedback and regularity in the lexicon. *Phonology* 24, 147–185.
[34] Wedel, A. 2012. Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition* 4, 319–355.
[35] Wedel, A, Jackson, S., Kaplan, A. 2013. Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech* 56, 395–417.
[36] Wedel, A., Kaplan, A., Jackson, J. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128, 179–186.
[37] Whitacre, J. M. 2010. Degeneracy: a link between evolvability, robustness and complexity in biological systems. *Theoretical Biology and Medical Modelling* 7, 1–17.
[38] Whitacre, J., Bender, A. 2010. Degeneracy: a design principle for achieving robustness and evolvability. *Journal of Theoretical Biology* 263, 143–153.
[39] Winter, B. 2014. Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. *BioEssays* 36, 960–967.
[40] Wright, R. 2004. A review of perceptual cues and robustness. In: Steriade, D., Kirchner, R., Hayes, B. (eds.), *Phonetically based phonology*. Cambridge, UK: Cambridge University Press, 34–57.
[41] Zuidema, W., de Boer, Bart. 2009. The evolution of combinatorial phonology. *Journal of Phonetics* 37, 125–144.

UNIVERSALITY IN CULTURAL TRANSMISSION

Bill Thompson

AI Lab, Vrije Universiteit Brussel
bill@ai.vub.ac.be

ABSTRACT

Many physical systems exhibit *universality*: system-level behaviour is invariant to differences in the micro-level interacting elements that make up the system. Here I explore the possibility that, broadly understood, this property may also be true of *cultural transmission*, the process through which languages gain and lose structure. I take a powerful but computationally expensive Bayesian model for unsupervised induction of phonetic categories – the *infinite mixture-of-Gaussians* model [3] – and adapt it to include a lightweight, psychologically plausible scheme for rapid approximate posterior inference. I use this model to simulate cultural transmission among populations of learners, and search for signatures of the insensitivity that underpins *universality*: I ask how the population-level distribution of phonetic categories varies – or not – as a function of the biases inherent in the model of category acquisition. I discuss the potential significance of this connection to physical systems, and argue that this principle of insensitivity may have interesting consequences for the evolution of phonetic systems.

Keywords: Cultural Transmission; Universality; Bayesian Inference; Phonetic Categories

1. INTRODUCTION

An astonishing number of physical systems can be said to exhibit the property of *universality*: macro-level behaviour of the system is invariant to variations in the details at the micro-level. For example, perhaps the most general case of universality lies in the fact that all physical systems in equilibrium are subject to the same laws of thermo-dynamics, regardless of the particular atomic interactions that underpin the system. Universality has been discovered in a remarkable range of contexts, from the critical exponents that describe phase-transitions in a range of molecularly diverse substances, to the balance of chaos and order in seemingly unconnected dynamical systems, ranging from hypothetical mathematical constructs like the Zeta function [7], to the unregulated bus networks of Cuernavaca, Mexico [5]. Universality has also been argued to be a principle of broad significance to multi-agent dynamical systems in AI [9], when understood to describe: "...any system of interacting elements whose qualitative or quantitative system-level behaviour includes characteristics that are invariant under changes in the individual behaviour and detailed interaction of the elements." [9, pp.2].

Here I adopt this perspective and explore the possibility that, in some important respects, cultural transmission may also exhibit dynamics reminiscent of universality. In particular, I suggest that under some conditions cultural transmission may exhibit universality with respect to characteristics of the inductive biases that underpin learning. In section 2, I describe a Bayesian model for unsupervised inference of phonetic categories that permits a flexible range of inductive biases. In section 3 I simulate cultural transmission under this model, and show that there are conditions where differences in these biases do not result in differences at the population-level distribution of phonetic categories. Finally, section 4 lays out some potential consequences for our understanding of the evolution of phonetic capacities.

2. A MODEL OF PHONETIC CATEGORY ACQUISITION

Here I lay out a Bayesian model of phonetic category learning grounded in statistical inference over distributional cues. Following e.g. [2], I adopt the *mixture of Gaussians* representation of phonetic categories: in particular, I adapt the *infinite mixture of Gaussians* (iMOG) model developed by [3], and reformulate important aspects to reflect a simple, lightweight algorithm for posterior inference.

Models of phonetic category learning based on distributional statistics have gained considerable attention recently [6], and have proven particularly useful tools for exploring difficult acquisition problems, such as the rich statistical dependencies that could allow lexical statistics to bootstrap phonetic category acquisition [3], or the latent hierarchical structure that allows generalisation between phonetic categories [8], for example. This trend is in line with the broader movement to explore domain-independent rational statistical inference as an explanation for inductive leaps that have traditionally been thought to indicate specialised inductive biases, particularly with respect to language e.g. [10]. Though the promise of these models is clear, there remain many fascinating open questions concerning the psychological mechanisms responsible for approximating the prohibitively complex computations that underpin these inferential models.

Here I chip away at the considerable computational resources required to implement unsupervised phonetic category acquisition in the iMOG while maintaining its desirable qualitative properties. In line with comparable variations on this model for of category acquisition *en général* [11], I simplify the model by implementing a psychologically plausible, sequential, greedy algorithm for approximating the posterior distribution over category assignments it implies, and by assuming a MAP point-estimate approach to parameter estimation for individual categories.

2.1. Phonetic Categories as Gaussian Distributions

Here I adopt the abstraction that a single acoustic feature f is relevant to the classification of speech sounds into categories. For example, f could be understood as voice onset time, or an absolute formant value, used in a language to distinguish one phonetic category from another. The learner observes a sequence of unlabelled (the learner does not know which category each sound represents) phonetic tokens $X = (x_1, \ldots, x_N)$, where each observation x_i is a speech sound exhibiting a value for f from a continuous range, for $i = 1 \ldots, N$. The model assumes that the learner's goal is to assign each observation x_i to a phonetic category c.

The major value of the iMOG is that it allows the number, C, of underlying phonetic categories to be inferred directly from the data, without being specified in advance or subject to an upper limit. Let $Z = (z_1, \ldots, z_n)$ be a partition on X, such that $z_i = c$ is an index which specifies that observation x_i has been assigned to category c. Each phonetic category c is assumed to be characterised by a Gaussian distribution with mean μ_c and variance σ_c^2. This allows a simple form for the likelihood that sound x_i would be produced from category c:

$$\Pr(x_i|\mu_c, \sigma_c^2) = \mathcal{N}(x_i; \mu_c, \sigma_c^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}. \quad (1)$$

The learner must infer the Gaussian parameters describing the distribution of sounds that will be produced from a category: she must induce estimates $\hat{\mu}_c$ and $\hat{\sigma}_c^2$. Again assuming Bayesian inference for these parameters, we must specify a prior $\Pr(\mu_c, \sigma_c^2)$ that captures the learner's initial expectations about μ_c and σ_c^2.

2.2. Prior Over Category Means and Variances

A natural conjugate prior for these parameters is the normal-inverse-chi-squared distribution. Dropping the index c for notational convenience, the prior is defined to be:

$$\begin{aligned}\Pr(\mu, \sigma^2|\Lambda) &= \mathcal{N}\mathcal{I}\chi^2(\mu, \sigma^2; \mu_0, k_0, v_0, \sigma_0^2) \\ &= \mathcal{N}(\mu; \mu_0, \sigma^2/k_0)\chi^{-2}(\sigma^2; v_0, \sigma_0^2) \\ &= \frac{1}{\beta}\sigma^{-1}(\sigma^2)^{-(v_0/2+1)}\exp(\gamma) \qquad (2)\end{aligned}$$

$$\beta = \frac{\sqrt{2\pi}}{\sqrt{k_0}}\Gamma(v_0/2)\left(\frac{2}{v_0\sigma_0^2}\right)^{v_0/2} \qquad (3)$$

$$\gamma = -\frac{1}{2\sigma^2}\left(v_0\sigma_0^2 + k_0[\mu_0 - \mu]^2\right), \qquad (4)$$

where $\Lambda = (\mu_0, k_0, \sigma_0^2, v_0)$ are the parameters of the prior. Though the expression for the prior is necessarily complex (it must dictate a probability density function over two independent continuous variables), this formulation affords some useful mathematical properties, and has an intuitive interpretation: μ_0 is the initial guess for μ_c, and k_0 defines the learner's confidence in that guess; σ_0^2 is the initial guess about the variance σ_c^2, and likewise v_0 determines the confidence in that guess.

2.3. Estimating μ_c and σ_c^2

Given a set of sounds $X_c = \{x_i\colon z_i = c\}$ believed to have been generated from category c, the learner combines this data with her prior beliefs to arrive at updated beliefs which are captured by the *posterior* distribution:

$$\Pr(\mu_c, \sigma_c^2|X_c, \Lambda) \sim \Pr(X_c|\mu_c, \sigma_c^2)\Pr(\mu_c, \sigma_c^2|\Lambda), \quad (5)$$

where the data likelihood factors into the product of the individual speech sounds:

$$\Pr(X_c|\mu_c, \sigma_c^2) = \prod_{x_i \in X_c} \Pr(x_i|\mu_c, \sigma_c^2). \qquad (6)$$

In contrast to [3], I assume that, for a given category, the learner induces estimates $\hat{\mu}_c$ and $\hat{\sigma}_c^2$ of the mean and variance of the Gaussian distribution for that category that reflect the *maximum a posteriori* (MAP) point-estimates, which maximise

$\Pr(\mu_c, \sigma_c^2|X_c, \Lambda)$: these are the Bayesian equivalent of maximum-likelihood estimates, and have known closed-form expressions that don't require heavy integration over the cumbersome 2-dimensional posterior:

$$\hat{\mu}_c = \frac{\mu_0 k_0 + N_c \bar{X}_c}{k_n} \tag{7}$$

$$\hat{\sigma}_c^2 = \frac{\nu_n \sigma_n^2}{\nu_n - 1}\,, \tag{8}$$

where $\bar{X}_c$ represents the sample mean of X_c; $N_c = |X_c|$ is the number of speech sounds assosciated with category c; $\nu_n = \nu_0 + N_c$; $k_n = k_0 + N_c$; and:

$$\sigma_n^2 = \frac{1}{\nu_n}\left(\nu_0\sigma_0^2 + \sum_{x_i \in X_c}(x_i - \bar{X}_c)^2 + \frac{N_c k_0}{N_c + k_0}[\mu_0 - \bar{X}_c]^2\right). \tag{9}$$

2.4. Prior Over Number of Categories

The iMOG assumes a Dirichlet process (DP) prior over possible partitions of the data into category assignments Z. The DP has two parameters: α, a *concentration parameter* which implements a bias to hypothesise more ($\alpha \to \infty$) or fewer ($\alpha \to 0$) underlying categories to explain the data; and G_0, a *base distribution* which in this case provides the prior over the parameters μ_c and σ_c^2 for individual categories $c = 1, \ldots, \infty$. Here, the normal-inverse-chi-squared distribution described in equation (2) acts as the base distribution over the kinds of categories learners are likely to encounter, and α is a parameter I will vary.

Full details of the complete statistical model can be found in [3]. However, here I leverage the fact that the DP can formulated as sequential process specifying a form for the categorisation decisions to be made upon encountering a new data-point. When deciding upon a category assignment for speech sound x_i, the learner is computing:

$$\Pr(z_i = c|x_i, Z_{\_i}) = \Pr(x_i|z_i = c, Z_{\_i})\Pr(z_i = c|Z_{\_i})\,, \tag{10}$$

where $Z_{\_i}$ represents the existing category assignments, and $\Pr(x_i|z_i = c, Z_{\_i}) = \Pr(x_i|\hat{\mu}_c, \hat{\sigma}_c^2)$ is the likelihood of observing speech sound x_i under this category given the sounds currently assigned to it, and the associated current estimate of its parameters, following eq. (1), with $\hat{\mu}_c$ and $\hat{\sigma}_c^2$ substituted for μ_c and σ_c^2 respectively. The sequential formulation of the DP, sometimes referred to as the *Chinese restaurant process*, allows us to formulate the prior over partitions into category assignments as a probabilistic choice between assigning a new observation to an existing category ($N_c \geq 1$) or creating a new category ($N_c = 0$):

$$\Pr(z_i = c|Z_{\_i}) = \begin{cases} N_c \Big/ \Big[\sum_{c=1}^{C}(N_c) - 1 + \alpha\Big], & \text{if } N_c \geq 1 \\ \alpha \Big/ \Big[\sum_{c=1}^{C}(N_c) - 1 + \alpha\Big], & \text{if } N_c = 0\,. \end{cases} \tag{11}$$

2.5. Rapid Sequential Approximate Posterior Inference

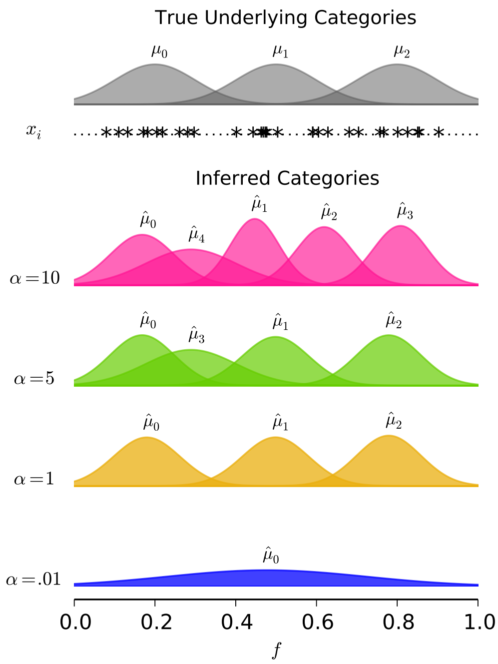
Finally, and perhaps most crucially, I assume a sequential, greedy updating scheme that provides a simple, psychologically plausible approximation to the posterior distribution over possible partitions Z. Specifically, I assume the learner arrives at a single fixed set of category assignments $\mathbf{Z}^*$ by sequentially choosing, upon receipt of each new observation x_i, to assign that data-point to whichever category maximises the posterior probability of assignment (eq. (10)) given the existing set of assignments, so that:

$$\mathbf{z_i^*} = \arg\max_c \Pr(z_i = c|x_i, Z_{\_i}) \tag{12}$$

The scheme begins by assigning the first observation to a new category $c = 1$, and cycles through further observations, in order, assigning each to a category c according to eq. (12), updating the estimates $\hat{\mu}_c$ and σ_c^2 for a given category's mean and variance as and when new observations are assigned. Equivalent updating schemes have been described in the machine-learning literature [12], and have been proposed for models of broader human categorisation behaviour [1, 11]. By deterministically choosing the MAP estimate for category assignment at each new data-point, the scheme allows extremely rapid inference, produces a single set of category assignments, and captures order-effects that are known to be characteristic of human cognition.

Figure 1 visualises an example of the model's inferences given noisy data generated from potentially overlapping phonetic categories. The figure shows the sets of categories inferred by the model, under four differing values for α, after observing $N = 30$ speech sounds sampled randomly with equal probability from three phonetic categories with parameters $\mu_1 = 0.2$, $\sigma_1^2 = 0.01$, $\mu_2 = 0.5$, $\sigma_2^2 = 0.01$, and $\mu_3 = 0.8$, $\sigma_3^2 = 0.01$. As is clear, in an individual learning scenario, the bias to infer more or fewer underlying categories (α) has a considerable impact on

Figure 1: Categories inferred by the model given $N = 30$ observed speech sounds (shown as stars on the dotted horizontal line) generated randomly from three underlying phonetic categories (top, grey), for four values of α. Prior parameters $\Lambda = (\mu_0 = 1., k_0 = 0.01 \sigma_0^2 = 0.01, \nu_0 = 2.)$ encode a vague prior that is essentially uniform across means, but weakly favours lower variance for individual categories.



the system inferred. At the extremes: higher values (e.g. $\alpha = 5.$, second row from the top, in red) cause the model to over-fit the data by inferring 5 distinct categories; lower values (e.g. $\alpha = 0.01$, bottom row, in blue) cause the model to discount variation in the data and hypothesise a single underlying category responsible for all observations.

3. RESULTS: CULTURAL TRANSMISSION OF PHONETIC CATEGORIES

Here I implement the model of phonetic category learning in a simple simulation of cultural transmission along a chain of learners. I assume categories are transmitted via *iterated learning* [4]: each learner observes a set of (unlabelled) speech sounds X^{t-1} generated from the phonetic categories of a previous learner, induces its own set of categories $\mathbf{Z}^{*\mathbf{t}}$, then uses these inferred categories to produce speech sounds X^t that form the observations of the next learner in the chain. The population-level characteristics of this process can be obtained by averaging over inferences made by learners along the full length of the chain.

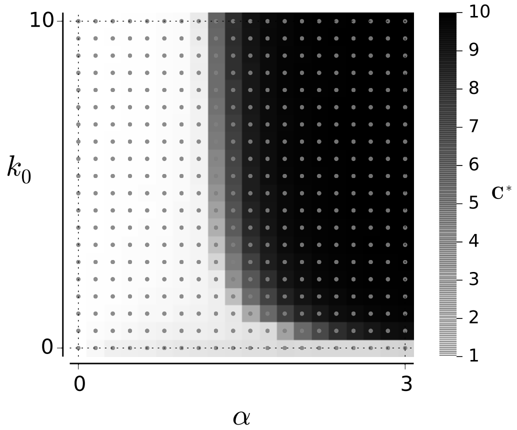
The hypothesis I set out in section 1 – that cultural transmission can be said to exhibit *universality* with respect to characteristics of individual biases – can be tested by exploring how these population-level characteristics vary – or not – as a function of the individual biases of learners in the chain. For example, here I focus on the average *number* of phonetic categories induced by learners along the chain, and how this varies with respect to the nature of two inductive biases: the concentration parameter α which dictates a bias to hypothesise more ($\alpha \to \infty$) or fewer ($\alpha \to 0$) distinct categories; and k_0, the *strength* of learners' bias in favour of particular values for the phonetic feature f. In particular, I analyse the model of transmission to ask how how these biases influence the likelihood of a phonemic merger: a collapsing of multiple distinct phonetic categories into one.

Each simulation was initialised with two distinct but overlapping phonetic categories ($\mu_1 = 0.5, \sigma_1 = 0.1, \mu_2 = 0.8, \sigma_2 = 0.1$), and run for 500 *generations*, or transmission episodes. Each learner observed $N = 10$ speech sounds generated randomly from the phonetic categories induced by the previous learner. All learners in the population share a bias to expect low variance in individual categories ($\sigma^0 = .01, \nu_0 = 10$), and a preference toward a particular range of speech sounds (an arbitrary value of f - this could represent a perceptual bias, for example), implemented by setting $\mu_0 = 0.6$. The strength of this preference for particular sounds is determined by the parameter $k_0 > 0$ (higher k_0 = stronger bias).

Figure 2 shows how the population-level average, $\mathbf{C}^*$, of the *number* of phonetic categories induced by learners varies as a function of k_0 and α. Each point in the grid represents a particular (k_0, α) pair, and the shade of the surrounding square gives $\mathbf{C}^*$, averaged over five replications of each simulation. Since the simulations were initialised with 2 phonetic categories, lighter shades ($\mathbf{C}^* < 2$) correspond to a phonetic merger, while darker shades ($\mathbf{C}^* > 2$) correspond to a *split*, or an increase in the number of categories.

While differences in α led to noticeably different inferences in *individual* learning (see figure 1), fine differences in α do not lead to fine differences in $\mathbf{C}^*$ over the course of cultural transmission constrained by a data bottleneck ($N = 10$). $\mathbf{C}^*$ does not appear to vary with the *strength* of α, only being sensitive to whether it is above or below a critical value. Roughly, any value of $\alpha < 1.5$ leads to a merger, while $\alpha > 1.5$ causes splits. While k_0 *can* play a role in determining $\mathbf{C}^*$ (there are some non-uniform

Figure 2: The population level average, $\mathbf{C}^*$, of the number of phonetic categories C induced by learners, as a function of individual biases α and k_0. $N = 10$, $\mu_0 = 0.6$, $\sigma_0^2 = 0.01$, $v_0 = 10$.



columns in the grid), it does not appear to do so under large regions of the parameter space. This insensitivity results from the modifications I made to the iMOG to allow a psychologically lightweight scheme for inference: MAP learning is known to lead to bias amplification over cultural transmission [4], and here brings about wide-scale bias-strength insensitivity.

4. DISCUSSION: TRANSMISSION, UNIVERSALITY, AND PHONETIC SYSTEMS

It is an intriguing possibility in its own right that, in at least some respects, cultural transmission of language can be likened to a broad class of physical systems via the concept of *universality*. The study of language transmission, and culture in general, may be enriched by exploring these connections further, in search of existing results relevant to dynamical systems that exhibit this kind of behaviour.

More specifically, universality in cultural transmission may have non-trivial consequences for understanding the origins of phonetic systems. For instance, if cultural transmission shapes language to match our phonetic biases, then universality implies that the process has more ammunition to work with: for example, pre-existing or domain-independent biases, even if extremely weak, could nevertheless be harnessed to shape phonetic systems.

This principle of insensitivity also implies an interesting asymmetry in our understanding of the relationship between phonetic biases and linguistic structure: for example, on the one hand, if we know a phonetic bias exists but do not know how significant a constraint it imposes, we could nevertheless make strong predictions about its population-level influence through cultural evolution; however, on the other hand, given only the population-level distributions of categories shown in figure 2, we could *not* make straightforward, reliable inferences about the *strength* of the underlying phonetic biases.

In general, wherever culture exhibits universality, we should be cautious of making direct inferences about cognition from the distributions of sounds we observe in a language.

5. SUMMARY

In this paper, I suggested that the concept of universality may be a useful – and in principle plausible – way to understand aspects of cultural transmission and the origins of our phonetic capacities. I took a recent, powerful model of unsupervised phonetic category induction, and reformulated some crucial assumptions to reflect a psychologically lightweight model of approximate inference. I simulated cultural transmission of phonetic categories under this revised model, and found conditions that suggest a dynamic reminiscent of universality in physical and multi-agent systems. While the results of these analyses reflect a very specific model, I hope to have demonstrated that the analogy is worthy of further investigation through more general means.

6. REFERENCES

[1] Anderson, J. R. 1991. The adaptive nature of human categorization. *Psychological Review* 98(3), 409–429.

[2] de Boer, B., Kuhl, P. K. Aug. 2003. Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online* 4(4), 129.

[3] Feldman, N. H., Griffiths, T. L., Goldwater, S., Morgan, J. L. Oct. 2013. A role for the developing lexicon in phonetic category acquisition. *Psychological review* 120(4), 751–78.

[4] Kirby, S., Dowman, M., Griffiths, T. L. 2007. Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences* 104(12), 5241–5245.

[5] Krbálek, M., Seba, P. July 2000. The statistical properties of the city transport in Cuernavaca (Mexico) and random matrix ensembles. *Journal of Physics A: Mathematical and General* 33(26), L229–L234.

[6] McMurray, B., Aslin, R. N., Toscano, J. C. Apr. 2009. Statistical learning of phonetic categories: insights from a computational approach. *Developmental science* 12(3), 369–78.

[7] Montgomery, H. L. 1973. The pair correlation of zeros of the zeta function. *Analytic Number The-*

ory, Proceedings of the 1972 St. Louis Symposiu. American Mathematical Society 181–193.

[8] Pajak, B., Bicknell, K., Levy, R. 2013. No A model of generalization in distributional learning of phonetic categories. *Proceedings of the 4th Workshop on Cognitive Modeling and Computational Linguistics* Sofia, Bulgaria.

[9] Parunak, H. V. D., Brueckner, S., Savit, R. 2004. Universality in Multi-Agent Systems. *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems* New York, New York, USA. 930–937.

[10] Perfors, A., Tenenbaum, J. B., Regier, T. Mar. 2011. The learnability of abstract syntactic principles. *Cognition* 118(3), 306–38.

[11] Sanborn, A. N., Griffiths, T. L., Navarro, D. J. Oct. 2010. Rational approximations to rational models: alternative algorithms for category learning. *Psychological review* 117(4), 1144–67.

[12] Wang, L., Dunson, D. B. Jan. 2011. Fast Bayesian Inference in Dirichlet Process Mixture Models. *Journal of computational and graphical statistics* 20(1).