# IS THERE A BIOLOGICAL GROUNDING OF PHONOLOGY? DETERMINING FACTORS, OPTIMIZATION, AND COMMUNICATIVE USAGE

*Susanne Fuchs* [1]*, Bernd Pompino-Marschall* [1,2] *& Pascal Perrier* [3]

[1] ZAS Centre for General Linguistics Berlin, [2] Humboldt University Berlin, [3] GIPSA-lab ICP, UMR CNRS 5009 & INP Grenoble

fuchs@zas.gwz-berlin.de, bernd.pompino-marschall@rz.hu-berlin.de, perrier@icp.inpg.fr

## ABSTRACT

This special session discusses the issue if there is a biological grounding of phonology. By reviewing current and past work from different speech research disciplines we suggest that (1) biological factors provide the limits, the frame of reference for phonology, (2) phonology is shaped by optimization processes taking into account the nonlinear relations of different representations of speech (acoustics, articulation, speech perception), and (3) sociolinguistic factors and communicative usage affect, for instance, speech acquisition and sound change. The first of the three suggestions is biological in nature whereas the last represents the non-biological nature of speech.

**Keywords:** physics, speech, biology of speech.

## DECLARATION

§1 All spoken languages are created equal […]

§n The free variation is one of the most precious of the rights of language. Every language may accordingly change, give rise to variance, using its many degrees of freedom, but shall be responsible for such abuses of this freedom as is defined by determining factors.

## 1. INTRODUCTION

"[In trying to understand] how human beings communicate by means of language, it is impossible for us to discount physical considerations, the facts of physics and physiology." (Halle 1954, cited in Ohala 1978, [18] p.5). It is hard to deny the biological foundation of speech since the speech signals we are able to realize and perceive stem from bio-physical systems, (1) the vocal apparatus and (2) the eyes and ears both controlled by the neural system. However, if all normal and fully developed human beings have more or less the same basic bio-physical system, why do we find such a variety of languages (at least 3000-6000 extant) with a tremendous amount of speech sounds and their combinations? How does the biological foundation of speech shape phonology?

Classical and recent approaches in phonology [25, 4, 5] have mainly worked on the definition of terminologies and methods to allow description and characterization of languages, including typological classification among them, and prediction of their potential diachronic evolution. However, this work consists in abstract formalizations of experimental observations of the languages' properties and variations. It does not provide a way to understand the origins of the different speech units, why some combinations of sounds do not exist, and why the possibilities of diachronic evolution are limited.

During the ICPhS conference 1983 in Utrecht a comparable issue was addressed in a session organized by Peter Ladefoged. Ladefoged [10] suggested that "sound patterns are the result of languages being a self-organizing social institution" (p.91). "Evolutionists teach us that such things are properties of a culture, and not of an individual's physiology" (p.94). Contrary to this view, Lindblom [12, 13] in the same session proposed that fundamental speech units and processes can be derived deductively from independent premises anchored in physiological and physical realities.

Since then, experimental techniques for studying speech production and perception have tremendously improved (among others Electromagnetic Articulography, Magnetic Resonance Imaging, Ultrasound, and Brain Imaging techniques). The increase of computer capabilities and the development of powerful software packages permitted the elaboration of complex realistic physiological models in order to study the underlying control mechanisms in the production and perception of sounds. Recently, a number of approaches have taken into account the characteristics of the speech production and perception apparatus and provided interesting suggestions that not only describe, but also explain some important aspects of the morphogenesis of language units [16, 21].

In phonology, on the other hand, e.g. constraint based models were developed during the last decade or so that ask for a natural grounding of universal principles [5, 20]. The lesson we should have learned from OT and related (phonological) models is that "universal", i.e. natural, constraints may shape phonological structure. Moreover, the interface model of Articulatory Phonology [2] lets us see the question of biological grounding of phonology in a new light. Consequently, we have decided to update this discussion in the light of recent findings and propose the special session entitled "Is there a biological grounding of phonology?". Our thoughts will be organized using three keywords: determining factors, optimization, and communicative usage.

## 2. DETERMINING FACTORS

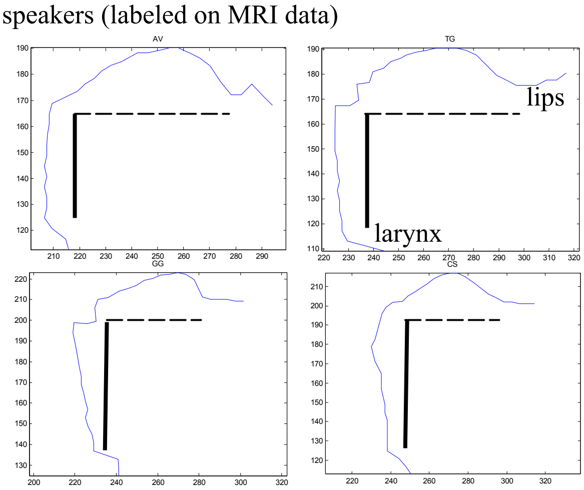
We suggest that the biological foundation of speech provides the determining factors for the development of phonology. Biology determines the limits, the frame of reference for phonology and specifies the course that languages can take in their development. Hereafter we will provide some examples of biological origin constraining the speech production process.

On the basis of tongue shapes in the mid-sagittal plane, Harshman et al. [9] have shown for English speaking subjects by a statistical procedure called PARAFAC, that two factors can explain most of the variance in the data. The 'front raising' factor is associated with movements along a high-front to low-back axis, and the 'back raising' factor is associated with movements along a low-front to high-back axis. Similar results were also reported in the literature for other languages (Icelandic, German, French) leading to the assumption that it could be a universal principle. But what would be the reason for its emergence?

Using a 2 D biomechanical model of the tongue for the analysis of a large amount of simulations, Fuchs & Perrier [7] proposed that these two factors would be a consequence of muscle anatomy and the internal structure of the tongue. They would not be speech specific. From this perspective, the front raising and back raising factors would be totally independent of any linguistic specification. Contrariwise, they seem to correspond to a biological constraint that influences the production of speech sounds, their variability, and their potential evolution.

In recent work Winkler et al. [27] we considered different vocal tract shapes as a biologically determining factor for individual articulatory control strategies aiming towards acoustic goal regions. Human adult vocal tracts differ particularly with respect to the length of the pharynx between males and females Fitch & Giedd [6]. Figure 1 displays the outer surface of different vocal tracts from 4 speaker's mid-sagittal MRI data. The plots on top exhibit larger horizontal proportions (schematically depicted as a dashed line) than vertical proportions (schematically depicted as a bold line) and the lower pair shows the reverse. As a starting point we suppose that speakers should have more space to move vertically in a tract with a large vertical space whereas they are more constrained (in terms of articulatory precision) in the horizontal direction. The opposite should be true for speakers with a larger horizontal than vertical vocal tract geometry. Consequently, to reach the same acoustic goal regions, speakers with different vocal tract shapes also need to posture their tongue differently, which in turn requires different muscle activation. Preliminary results of our experimental and modeling work provide evidence that in the production of French /i/ speakers with a large vertical vocal tract proportion show more upward than forward tongue movement. In our biomechanical simulations this movement was attributed to a higher activation of the Styloglossus.

**Figure 1:** Different outer vocal tract shapes from 4 speakers (labeled on MRI data)



Similarly, Brunner et al. [3] discussed the impact of the individual palate shape in the coronal plane as a determining factor for articulatory precision. Speakers with a more flat palate shape need to place their tongue more precisely than speakers with dome-shaped palates, since small articulatory changes in tongue height position have a larger impact on the area function and hence on the acoustics. By means of articulatory and acoustic data from 32 speakers we found that speakers with a flat palate shape realized a reduced token-to-token variability. Speakers with dome shaped palates were less constrained. Some had a large and some had a small degree of variability. Additionally, the actual effect of differences in the coronal palate shape was tested by means of a model. The influence of the palate shape on articulatory precision is generally strong for high vowels. However, this determining factor may be particularly relevant for language with a dense vowel inventory, but of less importance for the ones with a 3-vowel system.

## 3. OPTIMIZATION

Optimization processes are at work, taking into account the multimodal representations of speech (Perrier 2005). There are a number of concepts in the literature focusing on the optimization of the nonlinear relations of these representations.

### 3.1. Optimization between acoustic, perception and articulation

According to Stevens' [23, 24] 'Quantal nature of speech', spoken languages prefer those sounds or regions where articulatory changes have only little impact on the acoustic consequences. Following this perspective, articulatory regions of acoustic stability shape phonology.

Functional Phonology [1] and Hyper-& Hypo-speech [12] take into account the oppositional principles of articulatory economy and perceptual comprehension. According to them, the speech production systems in common with all motor systems, is organized to minimize effort, resulting in articulatory reduction, weakening or even deletion of movements. Minimization of effort, however, goes against the minimization of perceptual confusion within the communicative process. Functional Phonology as well as H&H theory focuses on the optimization between articulatory effort and perceptual recognition.

Another study discussing the relation between articulation and perception is reported in Liljencrant & Lindblom [11]. They hypothesized that a maximal perceptual distinctiveness principle underlies the distribution of the vowels. On this basis the authors could predict why /i/, /u/, and /a/ are present in almost all vowel systems.

The Dispersion-Focalization Theory (DFT) proposed by Schwartz et al. [22] also relies on an optimality criterion. It involves a maximal perceptual distinctiveness principle, but combines it with the concept of 'good perceptual objects', inspired from the 'Gestalt' theory. Schwartz et al. consider that 'focalized' frequency characteristics, i.e. spectra depicting a close proximity and even a merging of different maxima of energy, are perceptually more salient and more stable. Indeed, the DFT predicts very well the distribution of the vowels in the world languages as described in the UPSID data base [17].

### 3.2. Optimization at the motor control level

Comparing the production of alveolar stops with fricatives, Fuchs et al. [8] hypothesized differences in the control of these sounds: alveolar stops have a target above the contact location resulting in a collision of the tongue tip at the palate as opposed to a precise positioning of the tongue at the lateral margins of the palate for alveolar fricatives. The first assumption was based on Löfqvist and Gracco [15] who found for bilabial stops that articulatory gestures could be directed toward a target that is beyond the actual contact location. Thus, in the production of coronal stops the palate is used as a reference, which automatically blocks the tongue's movement at the required location in the vocal tract and guarantees the acoustic silent closure. In terms of stability and simplicity, such a control strategy seems to be extremely efficient in comparison to the control of a fine positioning. By means of kinematic data Fuchs et al. found strong support for the different control strategies. They concluded that the control strategy used in stop production would be more simple and stable than the one in fricative production. This difference in the complexity of control could be part of the explanation for the fact that most languages have more plosives than fricatives (Vallée et al.[26]).

## 4. COMMUNICATIVE USAGE

The question of the extent to which biological factors influence speech during the communicative usage is a very challenging one, since it raises many methodological problems, especially with disentangling linguistic, sociolinguistic, communicative, biological and frequency factors - to name just a few.

As we all know for a long time, in speech communication we not only transport meaning, but also signal our attitudes, emotions, values etc. [25]. And of course, spoken language in general would not exist without social communication. This point has been taken up recently by Exemplar Theory [19] modeling for instance speech acquisition as the development of a self-assembling system from variable environmental input.

Sociolinguistic variation in itself is one of the most important non-biological factors structuring spoken language as pointed out by Scobbie in this session.

## 5. OUTLOOK

So far the question of the biological grounding of phonology was mainly expressed as the question of how the internal structure of different phoneme inventories can be explained (cf. the above discussion of Dispersion Theory and DFT). And indeed, evidence for quite different perceptual factors determining these segmental inventories have been found. Some of the newer observations (cf. the paper of Flemming to this session) point in the direction that we should look also more closely at the perceptual cues for individual segments in the acoustics of connected speech – and not only from a purely concatenatively oriented point of view (e.g. sonority sequencing). Sonority sequencing itself is surely grounded in our productive and perceptual abilities. Aren't all speech signals prosodic at their core? The contributions of Slifka and Goldstein et al. to this session already point in this direction. In this same vein we might extend the H&H Theory to a model of continuous modulation within individual utterances. At the hierarchically structured prosodic domains we might look for specific hypothesised hyper-to-hypo contours giving rise to boundary markers. For example, downdrift and final lengthening might be considered biologically natural phenomena. Can we trace back phonological lengthening rules in phrasing (e.g. Chichewa penultimate lengthening) to such mechanisms?

Another question worth pursuing from the biological, i.e. anatomical, perspective seems to be the question of inter-speaker variability. Recently growing data on anatomical differences between individual vocal tracts in connection with the inter-speaker variability in articulatory movements observed for long times provoke some new questions. Do different vocal tracts require different neural motor control structures to yield the same categorical output? Or are their articulo-acoustic results taken care of by their listener as exemplars around a prototype? Until now, the discussion has centred around artificial prototypes (e.g. an acoustically defined 7-step height pattern of vowels in DFT). What would the picture be if we took into account productive constraints and anatomically induced exemplar variability? What can we learn from 'compensatory' articulation aiming at the same auditory target (cf. the paper by Honda et al in this session)?

## 6. ACKNOWLEDGEMENTS

This work was supported by grants from the German Research Council (DFG) GWZ 4/8-1 P1, PO 334/4-1 and by the P2R program of the French Foreign Office. Thanks to the members of the French-German network POPAART for reviewing several papers of this special session and Jim Scobbie for proof reading and helpful discussions.

## 7. REFERENCES

[1] Boersma, P. (1998) *Functional Phonology: Formalizing the interaction between articulatory and perceptual drives*. PhD dissertation. University of Amsterdam.

[2] Browman, C.P. & Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology* 6, 201-251.

[3] Brunner, J., Fuchs, S. & Perrier, P. (acc.) On the relation between the palate shape and articulatory behaviour. *JASA*.

[4] Chomsky, N. & Halle, M. (1968) *The Sound Patterns of English*. New York: Harper & Row.

[5] Clements, G.N. (1999) The geometry of phonological features (pp. 201-223). In J. Goldsmith (ed.) *Phonological Theory: The Essential Readings*. Oxford: Blackwell.

[6] Fitch, W.T. & Giedd, J. (1999) Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *JASA* 106(3): 1511-1522, 1999.

[7] Fuchs, S. & Perrier, P. (2005): On the complex nature of speech kinematics. *ZASPiL* 42: 137-165.

[8] Fuchs, S., Perrier, P., Geng, C. & Mooshammer, C. (2006). What role does the palate play in speech motor control? Insights from tongue kinematics for German alveolar obstruents. In J. Harrington & M. Tabain (eds.), *Speech Production: Models, Phonetic Processes, and Techniques* (pp. 149-164). Psychology Press: New-York.

[9] Harshman, R.A., Ladefoged, P.N. Goldstein, L. (1977) Factor analysis of tongue shapes. *JASA 62,* 693–707.

[10] Ladefoged, P. (1984) Out of the chaos comes order; Physical, biological, and structural patterns in phonetics. *Proc. of the Xth ICPhS Utrecht*, 83-95.
[11] Liljencrants, J. & Lindblom, B.E.F. (1972) Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language, 48*, 839-862.
[12] Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (ed.) *The production of speech* (pp. 217-245). New York: Springer.
[13] Lindblom, B. (1984) Can the models of evolutionary biology be applied to phonetic problems? *Proc. of the Xth ICPhS Utrecht*, 67-81.
[14] Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, W.J. & A. Marchal (eds.) *Speech Production and Speech Modelling* (pp. 403-439). Dordrecht: Kluver.
[15] Löfqvist, A. & Gracco, V.L. (1997) Lip and jaw kinematics in bilabial stop consonant production. *JSLHR* 40(4), 877-893.
[16] MacNeilage, P. F. (1998) The frame/content theory of evolution of speech production. *Behavioural and Brain Sciences* 21: 499-511.
[17] Maddieson, I. & Precoda, K. (1989) Updating UPSID, *UCLA WPP, 74,* 104-111.
[18] Ohala, J.J. (1978) Production of Tone. In Fromkin, V.A. (ed.) *Tone: A linguistic survey*. (pp. 5-39) NY, Academic Press.
[19] Pierrehumbert, J.B. (2006) The next toolkit. *J. Phonetics* 34, 516-530.
[20] Prince, A. & Smolensky, P. (2004) *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell.
[21] Schwartz, J.L., Abry, C., Boë, L.J. & Cathiard, M. (2002). Phonology in a theory of perception-for-action-control. In J. Durand & B. Laks (eds.) *Phonology: from Phonetics to Cognition* (pp. 255-280). Oxford: Oxford University Press.
[22] Schwartz, J.-L., Boë, L.-J., Vallée, N. & Abry, C. (1997) The Dispersion-Focalization Theory of vowel systems. *J. Phonetics 25(3)*, 255-286.
[23] Stevens, K.N. (1972) The quantal nature of speech: Evidence from articulatory-acoustic data. In P.B. Denes & J.R. Davis (eds.), *Human Communication: A Unified View* (pp.51-66). New York, McGraw-Hill.
[24] Stevens, K.N. (1989). On the quantal nature of speech. *J. Phonetics*, 17, 3-45.
[25] Trubetzkoy, N.(1939) Grundzüge der Phonologie. Prag.
[26] Vallée, N., Boë, L.J., Schwartz, J.L., Badin, P., & Abry, C. (2002) The weight of phonetic substance in the structure of sounds inventories. *ZASPiL, 28*, 145-168.
[27] Winkler, R., Fuchs, S. & Perrier, P. (2006) The relation between differences in vocal tract geometry and articulatory control strategies in the production of French vowels: Evidence from MRI and modelling. *Proc. of the ISSP Ubatuba*, 509-516.