Understanding speech production: The PILIOS approach

Susanne Fuchs, Pascal Perrier

Understanding speech production: The PILIOS[1] approach

Susanne Fuchs and Pascal Perrier
Understanding and modeling biological and physical mechanisms underlying speech production helps understanding speech motor control and its link to linguistic structure. Two of our recent studies are presented to illustrate this methodological idea. In the first study a realistic biomechanical tongue model is used to explore comprehensively the possible tongue shapes in the mid-sagittal plane. It is shown that the main directions of the tongue deformation observed in different languages are not related to any specific control, but rather to biomechanical and anatomical facts. The second study addresses the 'trough effect' observed in tongue shapes for bilabial consonants in VCV sequences. It is suggested that the combined analysis of kinematic and electromyographic data allows questioning previous interpretations of the trough effect with respect to speech motor control.

La compréhension et la modélisation des mécanismes biologiques et physiques sous-jacents à la production de la parole offrent un cadre fructueux pour comprendre le contrôle de la parole et ses liens avec le linguistique. Nous illustrons ce parti pris méthodologique avec deux études. Tout d'abord, nous avons exploité un modèle anthropomorphique de la langue pour étudier exhaustivement l'ensemble possible de ses formes dans le plan sagittal. Nous montrons que les directions principales de déformation observées dans différents langages ne sont pas imputables à un contrôle spécifique à la parole, mais à des caractéristiques anatomiques et biomécaniques intrinsèques. La seconde étude s'intéresse à l' 'effet de creux' observé sur la langue au cours de la consonne dans des séquences VCV. Nous suggérons que l'analyse simultanée de signaux EMG et de position associée à la prise en compte des mécanismes de génération des signaux EMG offre des alternatives crédibles aux interprétations classiques en termes de contrôle de la parole.

A. 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 cited in Ohala, 1978, p.5). It is hard to deny the biological foundation of speech, since the speech signals humans 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 capacities of the central nervous system. Now, if all normal and fully developed human beings share more or less the same basic biophysical 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? To which extent does the biological foundation of speech shape phonology?

Classical and recent approaches in phonology (Trubetzkoy, 1939; Chomsky and Halle, 1968; Clements 1999) 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. These studies consisted in abstract formalizations of limited experimental observations of the

[1] Beside being a Greek island, PILIOS is the abbreviation for "sPeech as the Interaction between LInguistics, cOgnition, and physicS", a currently funded project of our French-German research group.

languages' properties and variations. From our point of view they do not provide a convincing 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 constrained along certain directions.

There are at least two radically different views about what could be at the origin of sound patterns. Ladefoged (1984) suggested "*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). In this line is the endeavor in laboratory phonology to establish sociophonetics as a research area on its own (Pierrehumbert, 2006).

Contrary to the previous view, Lindblom (1983, 1984) proposed that fundamental speech units and processes could be derived deductively from independent premises anchored in physiological and physical realities. A number of investigations have comprehensively studied the characteristics of the speech production and perception apparatus. The findings of these studies provided the basis for some challenging theories that not only describe, but also explain important aspects of the morphogenesis of language units (MacNeilage, 1998; Schwartz *et al.,* 2002) or of phonological processes and sound change (Ohala, 1981).

However, it is not surprising that neither the pure biological nor the pure socio-cultural perspective offers on its own a comprehensive framework to study spoken languages and their diachronic evolution. Spoken language is the result of the combination of both factors, biology and culture (see Scobbie, 2007 for similar suggestions, but from a phonological perspective).

B. Combining biology and culture

We suggest that our biological system provides the frame of reference, *i.e.* what our articulators, eyes, ears, neural networks are able to do, but also, what goes beyond the human physical and cognitive capacities. The shape of the acoustic vowel space in the F1-F2 plane is a classical illustration for such a frame of reference (Boë *et al.*, 1989): high F2 values can never be associated with high F1 values due to the intrinsic characteristics of the vocal tract acoustics.

Another example is the 'separation' of the vocal tract into articulators that can be controlled relatively independently. This separation is the basis for the many degrees of freedom available to control the vocal tract shape, although not all of them are used in reality. In order to deal with the degrees of freedom and the nonlinear relations between movement (articulation), sound output (acoustics) and perception, various concepts have been proposed in the literature, focusing on the number of control parameters at different levels and on optimization principles using the many-to-one-relations between these levels. For instance, according to Stevens' (1989) '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.

In the same vein, Functional Phonology (Boersma, 1998) and Hyper-& Hypospeech (Lindblom, 1990) take into account the oppositional principles of articulatory economy and perceptual comprehension. According to them, the speech production system, like other motor systems, is organized to minimize effort, resulting in articulatory reduction, weakening or even deletion of movements. Minimization of effort, however, goes against perceptual discrimination within the communicative process. Both theories focus on the optimization between articulatory effort and perceptual recognition.

Another study discussing the relation between articulation and perception is reported in Liljencrants and Lindblom (1972). They hypothesized that a maximal perceptual distinctiveness principle underlies the distribution of the vowels. By means of this principle the authors were able to predict why /i/, /u/, and /a/ are present in almost all vowel systems.

The Dispersion-Focalization Theory (DFT) proposed by Schwartz *et al.* (1997) 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. The DFT predicts very well the distribution of the vowels in the world languages as described in the UPSID database (Maddieson & Precoda, 1989).

Moving within the physical frame of reference for speech production, using the relations between acoustic-articulation-perception in an optimal way, human beings transport not only meaning when communicating, but signal their attitudes, emotions, values *etc*. (Trubetzkoy, 1939). And of course, spoken language in general would not exist without social communication. A prototypical example supporting the role of cultural factors in sound change was recently brought up by Harrington (2006) who analyzed Queen Elisabeth's vowel production over the past 50 years. He was able to show a vowel shift and interpreted it as the Queen's wish not to "*sound old-fashioned on the one hand, but also still distinctly upper-class and very clearly differentiated from Estuary English on the other*".(p.454)

Over the last years our research focused mainly on the frame of reference in speech production. Therefore we will further concentrate on this point while acknowledging the importance of cultural and social factors.

C. Some methodological considerations

Our general approach has been to investigate the kinematic properties of various linguistic phenomena (phoneme distinctions, allophonic variations or prosodic influences) and to relate them to their corresponding acoustics. Our particular interest has been devoted to discover which of the relevant properties of the speech signals are controlled by the central nervous system and which ones are due to the properties of the speech production system (system-immanent). The questions we have been asking were: what is the impact of biomechanics, aerodynamics and vocal tract boundaries in shaping the kinematic output of a particular linguistic aspect? How much can be inferred from the kinematics about the underlying control mechanisms?

Such an approach has been possible by comparing data from human subjects with complex, physical models of the articulators and the vocal tract. It allowed us to specify certain physical parameters, and study their impact on articulatory movements and acoustics, thereby delivering not only descriptions about speech production, but also interpretations in terms of the underlying control. We believe that such a methodology

can prevent us from inferring ungrounded conclusions about the nature of speech production.
The basic underlying assumptions of our work are as follows:
1.) *Parsimony* in speech motor control: The complexity of speech signals and their time variations does not necessarily reflect the complexity of the underlying control. It is hypothesized that speech motor control is as parsimonious as possible while being efficient enough to ensure oral communication. A strategy to be parsimonious and efficient would be to build speech targets carrying semiotic information, which can be reached without controlling every little step towards them, but following the inherent properties of the speech production apparatus.
We certainly do not mean that speech production simply consists of the juxtaposition of physically determined transitions between targets. Successive targets and timing are specified via planning mechanisms that are more complex and require knowledge about the dynamics of the speech production apparatus. The complexity of speech motor control would be associated with the search for optimal ways to achieve and combine these goals.
Thus, it is assumed that: (1.) parts of the kinematic and acoustic characteristics are system-immanent, (2.) speech production properties largely contribute to shape perceptual goals, and (3.) the execution of speech gestures does not involve online complex cognitive processing, but planning does.
2.) The role of *feedback*: The durations of the transition intervals between units (phonemes or syllables) are very short in their majority. This duration prevents any kind of online use of auditory, proprioceptive or tactile long latency feedback from the cortex and limits the role of auditory feedback mainly to a belated correction or to the suprasegmental level (e.g. intensity, pitch). Consequently, no change of the central motor commands can be achieved during the execution of articulatory movements. The accuracy of articulatory movements is ensured thanks to non-cortical short delay feedback, which allows local corrections of muscles activations.
3.) Internal representations of the speech production apparatus: We adhere to the opinion that motor command sequences underlying the production of larger speech units are the result of an optimal planning. This planning is based on simulations of vocal tract behavior in the brain, thanks to the use of internal representations of the speech production apparatus, called 'internal models' (see Jordan and Rumelhart, 1992, Kawato, 1999). Recent works on brain imaging strongly support the existence of such internal models (Angelaki *et al.*, 2004) in the control of human movements. Consistent with the assumption of parsimony, it is assumed that speech production uses simple internal control models giving a rough account of the physical characteristics of the speech production apparatus (Perrier, 2006).
In this paper we won't be able to provide experimental support for all these points mentioned here, but we will concentrate on two examples of our own work that show the importance of the physical properties (biomechanics and aerodynamics) in speech production.

D. Two examples from our own work

D.1 Degrees of freedom in tongue movement

The production of speech requires the simultaneous control of at least thirty different muscles. However, at the same time classical articulatory descriptions of vowel production are limited to a small number of parameters such as high versus low, front versus back for the tongue, and rounded versus spread for the lips. Hence, the understanding of speech motor control requires a reduction of the dimensionality from the muscle control space to a more functional, speech-related control space. The dimensions of the functional, speech-related control space will hereafter be called the degrees of freedom of the vocal tract. The reduction in dimensionality shows to what extent specific muscles are coordinated to produce meaningful sounds in the different languages. Previous works on the degrees of freedom were mainly based on statistic analyses of kinematic data. Harshman *et al.* (1977), Jackson (1988), Nix *et al.* (1996) and Hoole (1999) applied a PARAFAC analysis to x-ray or EMA data for English, Icelandic, and German. Maeda (1990) ran a guided principal component analysis (PCA) on x-ray data of French.

Although four different languages with different vowel inventories were analyzed in these studies, most of the results presented show that more than 90% of the variance observed in the tongue shapes can be ascribed along two main degrees of freedom: (1) a movement of the tongue body along a high-front to low-back axis (called 'front raising' in Harshman *et al.*, 1977) and (2) a bunching of the tongue along a high-back to low-front axis (called 'back raising'). Jackson (1988) found that the number of degrees of freedom were language specific, i.e. different for English and Icelandic. However, his PARAFAC analysis was then proved to be degenerated by Nix *et al.* (1996), who reanalyzed the same data set.

The results of these studies lead to questions about the origin of the two main degrees of freedom: common to different languages – are they learned, speech-specific actions, or are they due to basic properties of the speech production mechanism? In the following we will explore the hypothesis that the two main degrees of freedom have their origin in the anatomical and biomechanical properties of the speech production apparatus. Toward this aim, a 2D biomechanical model of the tongue was used to generate a large set of tongue configurations, on which a PCA was run in order to extract the main axes of deformation.

First results were presented in Perrier *et al.* (2000). They were based on a gaussian sampling of the motor control space with the commands around the rest position as an average vector. These simulations were limited to the analysis of tongue configurations during vowel production, excluding those, which were too close to the palate. In this paper we propose an extension of the previous work, covering a very broad range of tongue shapes. We adopted a uniform sampling method and included tongue configurations in slight contacts with the palate. In doing so, our simulations cover the whole range of tongue shapes that can be generated by the model. Thus, 9000 tongue configurations were simulated and analyzed with the classical PCA procedure (see Perrier *et al*., 2000 for details). The results of the PCA are depicted in figure 1 with a variation of +/-1 standard deviation around the mean value along each of the principal axes. The first and second factors clearly correspond to the typical front and back raising patterns. The third factor can be associated with a vertical downward movement of the tongue body and the variability associated with the fourth factor is rather marginal.

In the majority of the studies based on statistical analyses of articulatory data, more than 90% of the variance observed for a subject was described by the first two factors, while in our study 3 factors are necessary to reach approximately the same level of description. Results are as follows: the first factor explains 69% of the variance, the first two factors 88%, the first three factors 96% and the first four factors 99%. The slightly greater number of factors is in agreement with Nix *et al.'s* (1996) findings, which showed that when the tongue shapes of 6 speakers were analyzed together, 4 factors were necessary to reach the same level of description in comparison to the 2 factors extracted from the data of a single subject. Since our data were generated from a variety of random muscle commands relevant for vowel production, they may be more general, analogous to the combined data from 6 speakers.

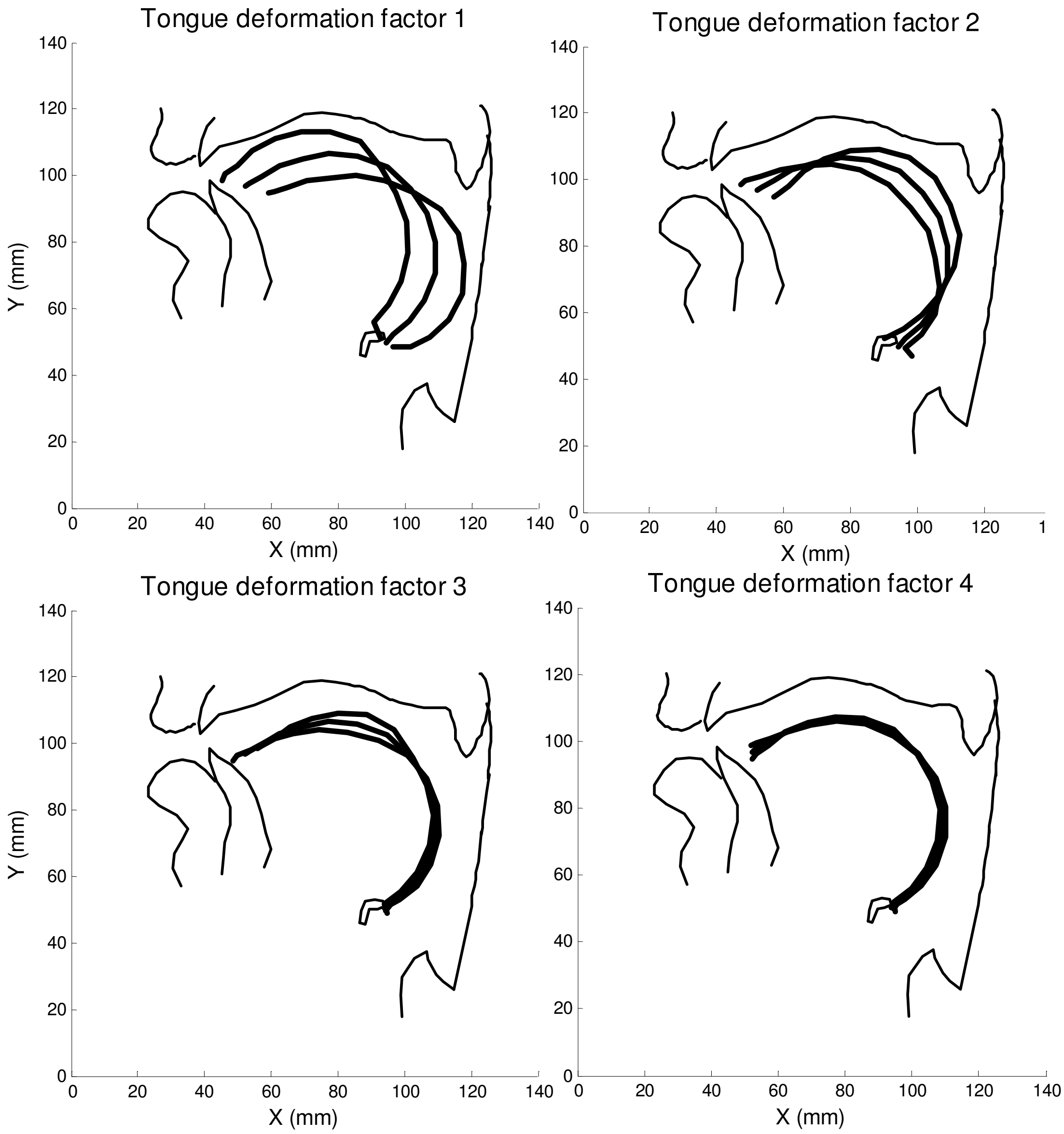

Figure 1: Tongue deformations with +/- 1 std. along the main principal axis for each factor. Dotted line: neutral position, solid line: positive deviation from the average, dashed-dotted line: negative deviation from the average

We conclude that the degrees of freedom in vowel production extracted from our simulations for French, and found in several studies for German, Icelandic, and English, are a result of the anatomical and biomechanical properties of the tongue and therefore not language-specific. Speech motor control uses these degrees of freedom to determine and differentiate speech articulations with respect to the various sounds of a language.

D.2 The trough effect

The trough effect (also called 'trough') can be described as a momentary deactivation of tongue movement during the consonant when it is a bilabial (Bell-Berti and Harris, 1974; Gay, 1975). It is a phenomenon occurring in a VCV-sequence, where both vowel phonemes are similar and the consonant is produced with an articulator which is assumed to be unspecified for vowel production, e.g. in /apa/. Discovering troughs was surprising, since one could expect the tongue to be activated from the first to the second vowel (e.g. aCa or iCi) and the bilabial closure to be realized independently of the surrounding vowels. Although the deactivation patterns and tongue lowering are very small in comparison to other kinematic events, several wide ranging theoretical issues were raised regarding this phenomenon, especially with respect to coarticulation, anticipation, and segment-by-segment activation.

The trough effect was taken as evidence that, at least at an electromyographic (EMG) level, anticipatory coarticulation is not taking place (Bell-Berti and Harris, 1974). Gay (1975) interpreted the trough as evidence against the ubiquity of anticipatory gestures occurring one segment ahead, without rejecting anticipation in general. In a recent paper, Lindblom *et al.* (2002) suggested that the trough effect would provide evidence for a segment-by-segment activation in opposition to Öhman's model of an underlying independent vowel cycle with a superimposed consonant gesture. The authors denied any potential aerodynamic explanations.

However, there is evidence in the literature for possible aerodynamic effects on the tongue surface. Svirsky *et al.* (1997) did not intend to study the trough effect, but investigated tissue compliance in the production of the voicing contrast. They used a comparable dataset, but included additionally /ama/ in their corpus. The results of their study provide evidence for a significantly larger tongue displacement, i.e. tongue lowering, in /b/ than in /p/, but no significant differences from zero were found for /m/. These results speak for an aerodynamic component, since tongue lowering coincides with these sounds where intraoral pressure is able to build up. However, since intraoral pressure is higher in /p/ than /b/, tongue lowering should go in the opposite direction (/p/>/b/) to the reported one (/b/>/p/). Hence, an active control strategy to increase the oral cavity in /b/ and maintain the transglottal pressure difference for the production of voicing was supposed to be involved too.

Based on these previous findings we are interested to answer the following question: Do troughs occur in VCV-sequences where C is a nasal? Can an aerodynamic explanation of the trough be eliminated?

Two experiments have been carried out in order to answer these questions (Fuchs *et al.* 2004). Work in progress will be presented here. In the first experiment 3 German speakers were recorded by means of hooked wire electromyography (EMG) and

acoustics. Activities of the following tongue muscles were registered for all subjects: Genioglossus posterior (GGP), Genioglossus anterior (GGA), and Styloglossus (SG). The speech material consisted of VCV-sequences where V was either /i/, /a/ or /u/ and C /p/, /b/ or /m/. Each target word was embedded in the carrier sentence: Habe X besucht (Visited X) and repeated up to 17 times. The post-processing procedure is described in detail in Fuchs *et al.* (2004). Ensemble averages for different muscle activities were lined up with the consonant onset, defined as the second formant offset or amplitude reduction of the acoustic envelope in the acoustic data.
All EMG results provide evidence for a muscular deactivation during the consonantal part of the VCV-sequence, *i.e.* troughs have their origin at a motor control level. However, the amplitude of the troughs differs among the subjects and muscles. Genioglossus posterior and Genioglossus anterior deactivation patterns were less pronounced in /m/ than in /p/ with an intermediate state for /b/ (for 2 out of 3 subjects in /i, u/-context, see figure 2 for an example). Results for styloglossus deactivation did not show such phonemic differences.

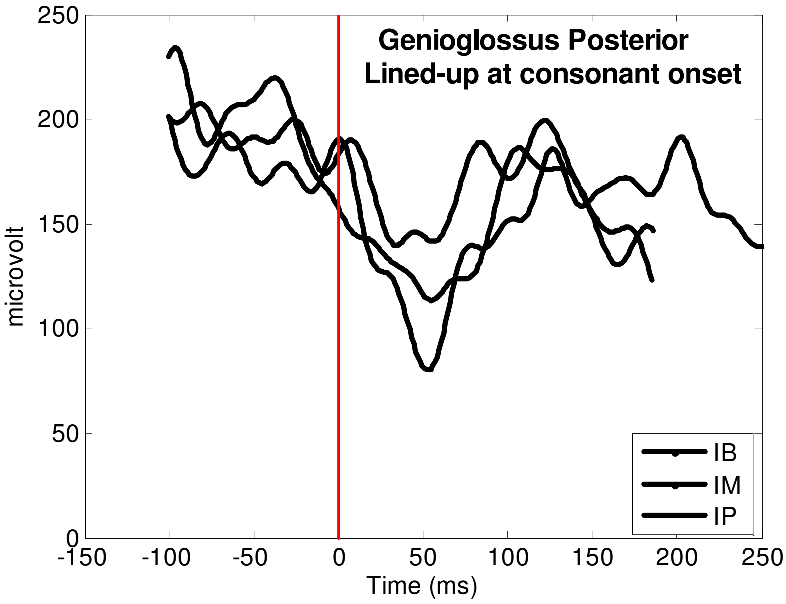

Figure 2: Averaged GGP activity lined-up at the acoustically defined consonant onset for CK /iCi/, dotted line correspond to C=/m/, dashed line to C=/b/, and solid line to C=/p/; vertical line = acoustically defined consonant onset

Note that electromyographic signals do not provide a direct mirror of the muscular control descending from the central nervous system. According to Feldman *et al.* (1990), EMG signals are a combination of central and feedback inputs. Changes in muscle length caused by external factors such as gravity or external forces can induce changes in EMG activity due to low-level feedback loops. On this basis, it is possible to consider that air pressure forces could modify tongue shapes and influence EMG activity. If muscle fibers/tongue position can be modified by an increase of intraoral pressure known for voiceless stop production, our findings can be interpreted with respect to aerodynamic constraints (see also Engstrand, 1981). If tongue lowering were actively

controlled, one would expect muscle activation for one muscle, but deactivation for the other muscle. However, a global deactivation pattern has been found in our data.
In order to relate the deactivation patterns to the kinematic output, an EMA (Electromagnetic Articulography) experiment was carried out with the same subjects, using the same speech material. The experiment consisted of two parts, one under normal speech condition, and one with a 5 mm bite block between the second molars. The latter was intended to fix the jaw and investigate the potential trade-off between tongue and jaw movements (see Vilain *et al.*, 2000). Three coils were attached to the tongue, one coil at the lower incisors (jaw), one at the vermillion border of the upper lip, one at the lower lip and two coils to compensate for helmet movements (one at the upper incisors and one at the bridge of the nose). All target words were repeated 19 times in the normal condition and 15 times in the bite block condition. Two successive gestures were labeled (the lower lip closing gesture from V1 to C: time points 1-7 in figure 3 and the opening gesture from C to V2: time points 8-14 in figure 3). So far only the data for two out of three speakers have been analyzed in the /aCa/ and /iCi/ context. We will focus on the /i/-context where comparable EMG data are available.
In figure 3 the average tongue blade contours are displayed from the beginning of lower lip closure to the following vowel (normal speech condition, /i/-context, same subject as in figure 2). Troughs are mostly pronounced in /ipi/ followed by /ibi/ and no changes in tongue position have been found during /imi/. These results are consistent across speakers and conditions in the /i/-context.

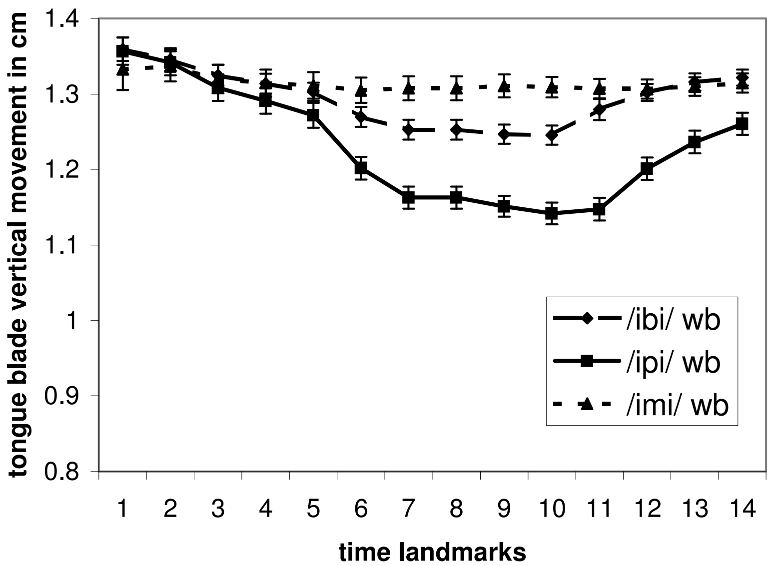

Figure 3: Averaged vertical tongue blade positions for /iCi/ (from time landmark 1-14) with +/-1 standard error for /ipi/= black solid line, /ibi/= dashed line, /imi/= dotted line; normal speech condition, speaker CK

Tongue blade lowering was never found in the /a/-context in the normal speech condition, and only to a very small extent in the bite block condition. Such vowel dependent results have also been reported by Vazquez-Alvarez and Hewlett (2007) using ultrasound.

Since our preliminary kinematic and electromyographic results are quite similar (largest troughs for /p/ followed by /b/, no troughs or very small ones for /m/), and we suppose intraoral air pressure differences in the same direction (highest pressure for /p/, lower pressure for /b/, atmospheric pressure for /m/), we conclude that aerodynamic factors are involved in the production of the trough effect. However, they may not be the only explanation, since troughs are also found in the nasal.

The trough effect is another example which shows that (1.) physical factors have a major impact on speech production characteristics, (2.) studying system-immanent factors may prevent us from inferring ungrounded conclusions about the nature of various speech production phenomena, (3.) not all speech production properties are controlled by the central nervous system, and (4.) ….." *it is impossible for us to discount physical considerations."* (Halle, cited in Ohala 1978, p.5).

Acknowledgments

We express our appreciation to all collaborators who were involved in developing this theoretical perspective over the years, in particular Bernd Pompino-Marschall who was involved in organizing a special session "Is there a biological grounding of phonology?" at the ICPhS in Saarbrücken, and Phil Hoole, Michiko Inoue, Emi Murano, Kyoshi Honda, Christian Geng, and Christian Kroos for their expertise and participation in the EMG experiment. This work was supported by a grant from the DFG, the BMBF, the CNRS, the French Ministry of Foreign Affairs within the P2R framework dedicated to the POPAART project, and the DFH Saarbrücken to the PILIOS project. It is dedicated to Dieter Fuchs.

References

Angelaki, D.E., Shaikh, A.G., Green, A.M. & Dickman, J.D. (2004). Neurons compute internal models of the physical laws of motion. *Letters to Nature* 430: 560-564.

Bell-Berti, F. & Harris, K. (1974). More on the motor organization of speech gestures. *Haskins Lab. Status Report on Speech Research* SR-37/38: 73-77.

Boë L.J., Perrier P., Guérin B. & Schwartz J.L. (1989). Maximal vowel space. *Proc. of the Eurospeech, Paris*, Vol. 2: 281-284.

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

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

Chomsky, N. & Halle, M. (1968). *The sound patterns of English*. New York: Harper & Row.

Clements, G.N. (1999). The geometry of phonological features. In J. Goldsmith (ed.) *Phonological theory: The essential readings*. Oxford: Blackwell: 201-223.

Engstrand, O. (1981). Acoustic constraints or invariant input representation? An experimental study of selected articulatory movements and targets. *Reports from Uppsala University Department of Linguistics* 7, 67-95.

Feldman A.G., Adamovich S.V., Ostry D.J. & Flanagan J.R. (1990). The origins of Electromyograms - explanations based on the Equilibrium Point Hypothesis. In J.W. Winters & S.L.Y. Woo (eds.): *Multiple muscle systems: biomechanics and movement organization*. Berlin: Springer: 195-213.

Fuchs, S., Hoole, P., Brunner, J. & Inoue, M. (2004). The trough effect: An aerodynamic

phenomenon? *CD-ROM From sound to sense*, MIT meeting.
Gay, T. (1975). Some electromyographic measures of coarticulation in VCV-utterances. *Haskins Lab. Status Report on Speech Research* SR-44: 137-145.
Harshman, R.A., Ladefoged, P.N. & Goldstein, L. (1977). Factor analysis of tongue shapes. *JASA* 62: 693–707.
Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *J. Phonetics* 34: 439–457.
Hoole, P. (1999). On the lingual organization of the German vowel system. *JASA* 106(2): 1020-1032.
Jackson, M.T.T. (1988). Analysis of tongue positions: Language-specific and cross-linguistic models. *JASA* 84(1): 124-143.
Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science*, 16, 316–354.
Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinions in Neurobiology*, *9*: 718–727.
Ladefoged, P. (1984). Out of the chaos comes order; Physical, biological, and structural patterns in phonetics. *Proc. of the Xth ICPhS Utrecht*: 83-95.
Liljencrants, J. & Lindblom, B.E.F. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language, 48*: 839-862.
Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (ed.) *The production of speech.* New York: Springer: 217-245.
Lindblom, B. (1984). Can the models of evolutionary biology be applied to phonetic problems? *Proc. of the Xth ICPhS Utrecht*: 67-81.
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* Dordrecht: Kluver: 403-439.
Lindblom, B., Sussman, H.M., Modarresi, G. & Burlingame, E. (2002). The trough effect: implications for speech motor programming. *Phonetica* 59 (4): 245-262.
MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioural and Brain Sciences* 21: 499-511.
Maeda, S. (1990). Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In W.J. Hardcastle & A. Marchal (eds.): *Speech production and speech modelling.* Dordrecht: Kluwer Academic Publishers: 131-150.
Maddieson, I. & Precoda, K. (1989). Updating UPSID, *UCLA WPP, 74:* 104-111.
Nix, D. A., Papcun, G., Hogden J., & Zlokarnik, I. (1996). Two cross-linguistic factors underlying tongue shapes for vowels. *JASA*: 3707-3717.
Ohala, J.J. (1978). Production of Tone. In Fromkin, V.A. (ed.) *Tone: A linguistic survey*. NY, Academic Press: 5-39.
Ohala, J.J. (1981). The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, & M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior.* Chicago: Chicago Ling. Soc.: 178 - 203.
Perrier, P., Perkell, J., Payan, Y., Zandipour, M., Guenther, F. & Khalighi, A. (2000). Degrees of freedom of tongue movements in speech may be constrained by biomechanics. *Proc. of the ISCLP Bejing*, 2: 162-165.
Perrier, P. (2006). About speech motor complexity. In J. Harrington & M. Tabain (eds.):

Towards a better understanding of speech production processes. Psychology Press: New York and Hove.
Pierrehumbert, J.B. (2006). The next toolkit. *J. Phonetics* 34: 516-530.
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.* Oxford: Oxford University Press: 255-280.
Schwartz, J.-L., Boë, L.-J., Vallée, N. & Abry, C. (1997). The Dispersion-Focalization Theory of vowel systems. *J. Phonetics* 25(3): 255-286.
Scobbie, J.M. (2007). Biological and social grouding of Phonology: Variation as a research tool. *Proc. of the ICPhS Saarbrücken*: 225-228
Stevens, K.N. (1989). On the quantal nature of speech. *J. Phonetics*, 17, 3-45.
Svirsky, M., Stevens, K., Matthies, M., Manzella, J., Perkell, J. & Wilhelms-Tricarico, R. (1997) Tongue surface displacement during bilabial stops. *JASA* 102: 562-571.
Trubetzkoy, N.(1939). *Grundzüge der Phonologie*. Prag.
Vazquez-Alvarez, Y. and Hewlett, N. (2007). The 'trough effect' : an ultrasound study. *Phonetica* 64(2-3): 105-121.
Vilain, A., Abry, C. and Badin, P. (2000). Coproduction strategies in French VCVS: Confronting Öhman's model with adult and developmental articulatory data. *Proc. of the 5th Seminar on Speech Production: Models and Data. Kloster Seeon*: 81-84.

Contact information:

Susanne Fuchs
ZAS/Phonetik
Schützenstrasse 18
10117 Berlin
Germany

Tel:++49 30 20192 569
Fax: ++49 30 20192 402
Email: fuchs@zas.gwz-berlin.de

Pascal Perrier
ICP/GIPSA-lab
961 rue de la Houille Blanche
Domaine universitaire, BP 46
38402 Saint Martin d'Hères cedex
France
Email: pascal.perrier@gipsa-lab.inpg.fr