

Comments on "The Equilibrium Point Hypothesis and its application to speech motor control" by P. Perrier, D.J. Ostry, and R. Laboissière (JSHR, Vol. 39, 365-378, April 1996)

TO APPEAR IN A SPECIAL ISSUE OF THE *BULLETIN DE LA COMMUNICATION PARLEE*

The EPH and control spaces for relaxation movements

or

«When is movement actually needed to control movement ?»

Christian Abry, Pierre Badin, Khaled Mawass & Xavier Pelorson

Institut de la Communication Parlée
UPRESA CNRS 5009, INPG – Université Stendhal
46, Av. Félix Viallet, F-38031 Grenoble Cedex 01, France

Introduction

In this commentary, we will address three main questions, all three concerning the *generation* and *control* of movement in speech, and more specifically the *representations of the dynamics* of its control parameters. For space sake, we will not recall experiments in favour of *static* representations of speech in *memory*, that we have reported in Abry & Badin (1996), concerning movement representation from *perception* with the issue: “Does movement in the ears and the eyes mean movement in the mind ?”

1. Control space for voicing and trills

First, is it viable to conceive of a speech sound generation system which produces movement without any “movement” of the control parameters ? One can anticipate that the answer is “yes”, as exemplified by plain phonation (normal voicing) and trill production. Control parameters for these productions are well known and less speculative than the EP control.

The literature on aero-biomechanical models of vocal folds vibration is dense (cf. e.g. Pelorson et al., 1995; Laboissière & Pelorson, 1995; Liljencrants, 1991; Ishizaka & Flanagan, 1972). It seems to be a widely accepted fact that vocal folds vibration is mainly controlled by three parameters: (i) one aerodynamic parameter, subglottal pressure, (ii) one mechanical parameter, vocal fold length, or vocal fold tension or the mass-tension ratio, and (iii) one geometrical parameter, vocal folds adduction or vocal folds distance at rest or glottis area at rest. The control space defined by these three parameters can be divided in two regions: one gives rise to self-sustained pseudo-periodic movements of the vocal folds for steady-state control parameters, while the other is a “no movement” region, based on equilibrium analysis. Laboissière & Pelorson (1995) give an example of these boundaries for a two-mass model of the vocal folds (cf. Fig. 1): in the space below the curves, the subglottal pressure is too low to allow vocal fold oscillations, while above the curves, oscillations can occur. Note that vocal fold vibration does not need any dynamic control of the three parameters.

Fig. 1: Example of control space for a two-mass model of vocal folds, for four different acoustics loads. The region below the curves is the “no vocal fold oscillation” region of the [subglottal pressure P_s / vocal fold stiffness k_1] control space, while the region above is the domain of voicing (from Laboissière & Pelorson, 1995).

The interest of this type of control space is to make available different articulatory-acoustic viable objects for different languages. Moreover, they must allow to explain sound changes corresponding to shifts from one stable state to another. In this spirit, one could argue that the great French “R” shift from a coronal articulation [r] to a dorsal one [R] could be accounted for in the following way. Remember first that French, like many other Romance languages, inherited from Latin a contrast between the “simple” flapped coronal [ʀ] and the “multiple” trilled coronal [r]. The study of historical documents and dialectal variations (Abry, 1977) has revealed that this contrast actually shifted from the [ʀ] vs. [r] opposition to the [ʀ] vs. [R] opposition, i.e. a shift of location of the trilled phoneme. Is it relevant to state that the control of a “simple” flapped uvular [R] is rather difficult ? Hence, the contrast could be as well enhanced by placing the flapped version at this dorsal location, but without recovering the stability of the product. Needless to say that the solution of using the flapped vs. trilled contrast in the uvular region had no known linguistic success. Finally, once the distance was enhanced, the trilled nature of the dorsal [R] could be abandoned, hence the change into a simple fricative [ʀ̂]. For a detailed study of this process in a dialectal area, see Abry (1977). Note that the analysis and the simulation of vowel systems using such *distance* and *stability* principles has resulted in the Dispersion Focalisation Theory (DFT) (Schwartz et al., 1997a, b); but as concerns trills, in spite of the first simulations mentioned above, this claim needs more computational evidence. In fact, it seems that modelling is far more advanced for snoring than for speech production, this being perhaps due to the recent medicalisation of the sleep apnea disease.

2. Control space for voicing vs. frication

Having described a first example of control space, which could stem plausibly from vocalizations and "raspberries" produced by infants, our second question will concern the viability of extrapolating such well known parameters to fricative production, which is mastered later. The answer is of course "yes" for sustained sounds. Thus we can present another example of control space, where the question is still whether consonants can be obtained without movement in the control parameters.

The importance of glottis and oral constriction coordination for the acoustic excitation sources in the vocal tract has been widely recognised and discussed in the literature (e.g. Stevens, 1971; Scully, 1971; Badin et al., 1996). The aerodynamic behaviour of the vocal tract is represented here by a simplified aerodynamic model (Badin et al., 1996). The tract is then viewed as two lumped constrictions between which the subglottal pressure is shared out. Although some refinements could be made, the discussion below is based on steady state considerations, for the sake of simplicity. Further, viscous losses are neglected, whereas the flow downstream the constrictions is considered to be fully turbulent, and thus no pressure recovery is considered. From a fluid mechanical point of view, this description is, of course, highly disputable: the flow through the glottis during phonation is not steady and the turbulent flow through a constriction during a fricative is essentially dominated by unsteady and viscous effects. However, it is expected that the major conclusions drawn below remain correct, even under such crude assumptions.

Denoting by A_g the low frequency component of glottis area, and by A_c the oral tract minimum constriction area, the pressure drops at the glottis ΔP_g and at the constriction ΔP_c are thus ruled by the following equations:

$$P_s = \Delta P_g + \Delta P_c, \text{ with } \Delta P_c = \frac{\rho}{2} \frac{U^2}{A_c^2} \text{ and } \Delta P_g = \frac{\rho}{2} \frac{U^2}{A_g^2}$$

where P_s is the subglottal pressure, ρ the air density, and U the constant volume flow velocity. It is also known that the amplitude of voicing increases with ΔP_g and reaches a maximum for a given A_g depending on ΔP_g (Stevens, 1971), while the amplitude of the friction noise source is proportional to $\Delta P_c P_s \Delta A_c^4$ (Badin et al., 1995).

It follows that, for a given subglottal pressure, increasing simultaneously voicing and frication amplitudes is contradictory: a balance must be found for voiced fricative consonants. This has been exemplified by simulations performed with our vocal tract model, that includes the simplified aerodynamical model, a two-mass model of the vocal fold, a friction noise source, and an acoustical reflection line analogue that takes care of source tract interaction (Badin et al., 1996). Both A_g and A_c were systematically varied in order to explore the $[A_g/A_c]$ control space: for each $[A_g/A_c]$ pair, the level of sound pressure radiated from a highly simplified fricative articulation excited by the voice excitation L_v only and by the frication excitation L_f only were computed. The requirement for a voiced fricative being the presence of both friction and voicing, the ratio between L_v and L_f was computed as a function of A_g and A_c . Fig. 2 displays, in grey levels, the quantity $-20|\log(L_v/L_f)|$ that represents the balance between voice and noise levels: when both levels are identical, this quantity is zero, and thus displayed in white, whereas when the difference increases it is more and more negative, coded in black below -12 dB. The narrow region which is not black represents thus the region of the control space that produces a signal for which the difference between the voice and noise components does not exceed 12 dB. It appears clearly that a rather strict coordination between the glottis and the oral constriction is needed to produce acceptable voiced fricatives (cf. also Mawass et al., 1997).

At this point, our questions to the authors could be: (i) How can you find the two EPs in such a control space ? (ii) Moreover, do these EPs speak to each other ? (iii) By the way, when they drive their system, how do they manage to “land” without any bumps ?

Anyway, it remains that, for articulator setting – i.e. onset, hold, and offset phases – we still need something equivalent to changes in control parameters, say EP movements or any other command making the articulators move. In fact, this would seem unavoidable both for voicing, trills and fricatives.

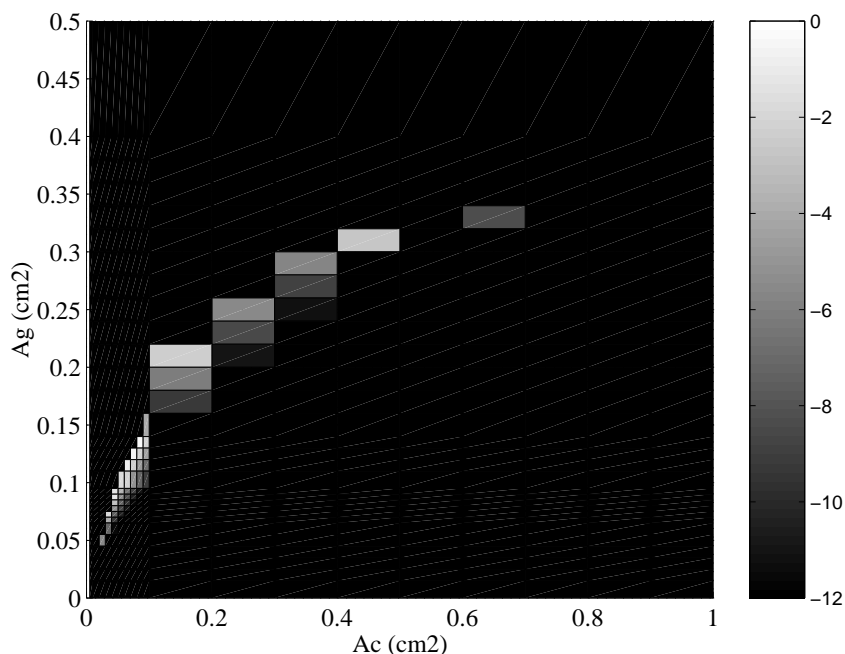


Fig. 2: Difference between voicing and friction amplitudes (in grey levels) in the (A_g/A_c) control space.

3. Which neural control for babbling and beyond ?

Finally, we will challenge this seemingly obvious answer to our previous question by asking for the possibility of a more primitive speech system which could also produce movement without changing the dynamics of control parameters.

We refer to canonical babbling onset, at about 7 months. In this case, contrary to e.g. trill production, the control parameters are fully unknown, roughly as for EP in the “non-spinal human” for speech. Thus it is fully a metaphor to speak about a *central pattern generator*, and of course of relaxation movements, in the human brain for speech - contrary to mastication in the brain stem rabbit (Luschei & Goldberg, 1981; for a comparison between locomotion, respiration and mastication, see Rossignol et al., 1988).

Consequently, remaining at the very same level of uncertainty - as concerns both the target paper and our comment -, we set forth the conception of babbling control generation propounded by MacNeilage and Davis (1990) and MacNeilage (to appear). The novelty of this proposal stays in the late recruitment of Broca’s area for speech. Arguments of MacNeilage and Davis (1990, p. 56) are manifold: (i) comparative neuroanatomy between monkeys and humans shows that the area *homologous* to Broca’s in monkeys plays no role in vocalisation control (Jürgens et al., 1982); (ii) in human cortical mapping Supplementary Motor Area is the only site evoking rhythmic syllable-like vocalisation (Penfield & Welch, 1951); (iii) irritative lesions of the SMA also produce such “babbling” movements (reports from Jonas, 1981; Poeck et al., 1984, ignore these data, and are

themselves ignored by MacNeilage; but they provide very similar descriptions, unfortunately without proper scanning of the SMA region [Abry et al., to appear]).

These arguments can be reinforced by two reviews on dendritic growth. According to Greenfield (1991): “the connections from Broca’s to the orofacial motor cortex develop at 12 to 15 months of age (Simonds & Scheibel’s [1991] neuroanatomical data, p. 582)”. In the same vein, Kent and Miolo (1995, p.308-309) notes that “one interpretation of these data [Scheibel (1993)] is that babbling reflects maturation of the motor cortex more than of Broca’s area”. There is thus a convergence to *discard for babbling emergence* the classical language production area.

Following this view, we provocatively conclude that humans until 15 months do not need any ramp-like EP changes, or any other command changes making the articulators move, since they do not recruit the system in charge of such controls, but a more primitive one, i.e. the SMA or a specific part of it (Tanji, 1996).

Recently, MacNeilage (to appear) took into account the discovery of so-called “mirror neurons” by Rizzolatti and colleagues in monkeys (Di Pellegrino et al., 1992), and in humans (Rizzolatti et al., 1996). Such neurons, located in Broca’s area, more specifically in Brodmann area 45 (and in the homologous area in monkeys) were selectively responsive both to execution and observation of *actions*. Hence, MacNeilage argues that specific language learnability could be mediated through this “lateral premotor system” (following Goldberg’s terminology, 1985), which receives inputs from temporal and posterior cortices, a system responsive to external simulation, available for learning by imitation. While the “medial premotor system”, i.e. the SMA, would be the basis for spontaneously generated action, namely babbling. This view answers both the need for a special speech module and a language specific acquisition device, thus meeting Chomsky’s programme of principles and parameters. This language specific domain could become the realm of EP changes, once sufficient decoupling from the jaw carrier of the end-effector articulators is achieved during the control maturation process.

Caveat

Even if MacNeilage were right, the frame that he proposes seems efficient essentially for plosive and approximant production, not for avoiding transient problems in trills and fricatives setting. Thus, our final understatement could be that we are still obviously unable to account for speech babbling and speech fricative production at the level of expertise attained both for mastication and voluntary (fine) biting (Luschei & Goldberg, 1981).

References

- Abry, C. (1977). Distribution contextuelle et géographique de deux vibrantes phonologiques dans quelques parlers francoprovençaux. *Revue de Linguistique Romane*, n° 163-164, (Tome 41), 241-284.
- Abry, C., & Badin, P. (1996). *Speech Mapping* as a framework for an integrated approach to the sensori-motor foundations of language. In *Proceedings of the 4th Speech Production Seminar - 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: from Control Strategies to Acoustics* (pp. 175-184). Autrans, France.
- Abry, C., Boë, L.J., Laboissière, R., & Schwartz, J.L. (to appear). Peer comment to “The frame/content theory of evolution of speech production”. *Behavioral and Brain Sciences*, forthcoming.
- Badin, P., Mawass, K., and Castelli, E. (1995). A model of frication noise source based on data from fricative consonants in vowel context. In Elenius, K. and Branderud, P. (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences* (Vol. 2, pp. 202-205). Stockholm, Sweden: Arne Strömbergs Grafiska Press.

- Badin, P., Mawass, K., Bailly, G., Vescovi, C., Beutemps, D., & Pelorson, X. (1996). Articulatory synthesis of fricative consonants : data and models. In *Proceedings of the 4th Speech Production Seminar - 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: from Control Strategies to Acoustics* (pp. 221-224). Aufrans, France.
- Crystal, D. (1997). *Dictionary of Linguistics and Phonetics*. Blackwell: Oxford.
- Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91, 176-180.
- Goldberg, G. (1985). Supplementary motor area structure and function: Review and hypothesis. *Behavioral and Brain Sciences*, 8, 567-616.
- Greenfield, P.M. (1991) Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behaviour. *Behavioral and Brain Sciences*, 14, 531-595.
- Ishizaka, K., & Flanagan, J.L. (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Systems Technical Journal*, 51, 1233-1268.
- Jonas, S. (1981). The supplementary motor area and speech emission. *Journal of Communication Disorders*, 14, 349-373.
- Jürgens, U., Kirsinger, A., & Crammon, D. von (1982). The effects of deep reaching lesions in the cortical face area on phonation: A combined case report and experimental study. *Cortex*, 18, 125-140.
- Kent, R.D., & Miolo, G. (1995). Phonetic abilities in the first year of life. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of Child Language* (pp. 303-334). Oxford: Blackwell.
- Laboissière, R., & Pelorson, X. (1995). Stability and bifurcations of the two-mass model oscillation: analysis of fluid mechanics effects and acoustical loading. In Elenius, K. and Branderud, P. (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 3, pp. 190-193). Stockholm, Sweden: Arne Strömbergs Grafiska Press.
- Liljencrants, J. (1991). A translating and rotating mass model of the vocal folds. *STL-QPSR*, 1/1991, 1-18.
- Luschei, E.S., & Goldberg, L.J. (1981). Neural mechanisms of mandibular control: mastication and voluntary biting. In *Handbook of physiology, The nervous system Vol. 2*. Washington, D.C.: American Physiological Soc.
- MacNeilage, P.F., & Davis, B.L. (1990). Acquisition of speech production: The achievement of segmental independence. In W.J. Hardcastle & A. Marchal (Eds.) *Speech Production and Speech Modelling* (pp. 55-68). Dordrecht: Kluwer.
- MacNeilage, P.F., (to appear). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*.
- Maddieson, I. (1989). Aerodynamic constraints on sound change: the case of bilabial trills. *UCLA Working Papers in Phonetics* 72: 91-115.
- Mawass, K., Badin, P., and Bailly, G. (1997). Synthesis of fricative consonants by audiovisual-to-articulatory inversion. To be published in *Proceedings of the 3rd EuroSpeech Conference* . Rhodos, Greece.
- McGowan, R.S. (1992). Tongue-tip trills and vocal-tract wall compliance. *Journal of the Acoustical Society of America*, 91, 2903-2910.
- Pelorson, X., Hirschberg, A., Wijnands, A.P.J., & Bailliet, H. (1995). Description of the flow through the vocal cords during phonation. *Acta Acustica*, 3, 191-202.
- Pelorson, X., Lallouache, T., Tournet, S., Bouffartigue, C., and Badin, P. (1994). Modeling the production of bilabial plosives: aerodynamical, geometrical and mechanical aspects. In *Proceedings of the International Conference on Spoken Language Processing* (Vol. 2, pp. 599-602). Yokohama, Japan.
- Penfield, W., & Welch, K. (1951). The supplementary motor area of the cerebral cortex. *A.M.A. Archives of Neurology and Psychiatry*, 66, 289-317.

- Poeck, K., de Bleser, R., & von Kayserlingk, D. (1984). Neurolinguistic status and localization of lesion in aphasic patients with exclusively consonant-vowel recurring utterances. *Brain*, *107*, 199-217.
- Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Perani, D., & Fazio, F. (1996). Localization of cortical areas responsive to the observation of hand grasping movements in humans: a PET study. *Experimental Brain Research*, *111*, 246-252.
- Rossignol, S., Lund, J.P., & Drew, T. (1988). The role of sensory inputs in regulating patterns of rhythmical movements in higher vertebrates: a comparison between locomotion, respiration and mastication. In A. Cohen, S. Rossignol, & S. Grillner (Eds.) *Neurocontrol of Rhythmic Movements in Vertebrates*. Wiley and Sons.
- Scheibel, A. (1993). Dendritic structure and language development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.) *Developmental Neurocognition, Speech and Face Processing in the First Year of Life* (pp. 51-62). Dordrecht: Kluwer.
- Schwartz, J.L., Boë, L.J., Vallée, N., & Abry C. (1997a). Major trends in vowel system inventories. *Journal of Phonetics*, *25*, 233-253.
- Schwartz, J.L., Boë, L.J., Vallée, N., & Abry C. (1997b). The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics*, *25*, 255-286.
- Scully, C. (1971). A comparison of /s/ and /z/ for an English speaker. *Language and Speech*, *14*, 187-200.
- Scully, C., Castelli, E., Brearley, E., and Shirt, M. (1992). Analysis and simulation of a speaker's aerodynamic and acoustic patterns for fricatives. *Journal of Phonetics*, *20*, 39-51.
- Simonds, R. and Scheibel, A. (1989). The postnatal development of the motor speech area. A preliminary study. *Brain and Language*, *37*, 42-58.
- Stevens, K.N. (1971). Airflow and turbulence noise for fricative and stop consonants: static considerations. *Journal of the Acoustical Society of America*, *50*, 1180-1192.
- Tanji, J. (1996). New concepts of the supplementary motor area. *Current Opinion in Neurobiology*, *6*, 782-787.