

Integrate, yes, but what and how? A computational approach of perceptuo-motor fusion in speech perception

Raphaël Laurent, Clément Moulin-Frier, Pierre Bessière, Jean-Luc Schwartz, Julien Diard

▶ **To cite this version:**

Raphaël Laurent, Clément Moulin-Frier, Pierre Bessière, Jean-Luc Schwartz, Julien Diard. Integrate, yes, but what and how? A computational approach of perceptuo-motor fusion in speech perception. Behavioral and Brain Sciences, 2013, 36 (4), pp.36-37. hal-00765973

HAL Id: hal-00765973
https://hal.science/hal-00765973

Submitted on 17 Dec 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

<Commentary Laurent et al. on Pickering & Garrod [BBS 35(3), 2012]>

<CT>Integrate, yes, but *what* and *how*? A computational approach of perceptuo-motor fusion in speech perception

<CA>Raphaël Laurent,[a,b] Clément Moulin-Frier,[a,c] Pierre Bessière,[b,d] Jean-Luc Schwartz,[a] and Julien Diard[e]

<CAA>[a]*GIPSA-Lab - CNRS UMR 5216, 11 rue des Mathématiques, Grenoble University, Grenoble Campus BP46, F-38402 Saint Martin D'Hères Cedex, France;* [b]*e-Motion team - INRIA Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France;* [c]*FLOWERS team - INRIA Bordeaux Sud-Ouest, 200 avenue de la Vieille Tour, 33405 Talence Cedex, France;* [d]*Laboratoire de Physiologie de la Perception et de l'Action - CNRS UMR 7152, Collège de France, Paris, France;* [e]*Laboratoire de Psychologie et NeuroCognition, CNRS UMR 5105, Grenoble University, Université Pierre Mendès France BSHM, BP47, 38040 Grenoble Cedex 9, France.*

Raphael.Laurent@gipsa-lab.grenoble-inp.fr
http://www.gipsa-lab.grenoble-inp.fr/page_pro.php?vid=1238
clement.moulin-frier@inria.fr
http://www.gipsa-lab.grenoble-inp.fr/~clement.moulin-frier/cv_en.html
Pierre.Bessiere@College-de-France.fr
http://www.Bayesian-Programming.org
Jean-Luc.Schwartz@gipsa-lab.grenoble-inp.fr
http://www.gipsa-lab.grenoble-inp.fr/~jean-luc.schwartz/
Julien.Diard@upmf-grenoble.fr
http://diard.wordpress.com/

<C-AB>
Abstract: We consider a computational model comparing the possible roles of "association" and "simulation" in phonetic decoding, demonstrating that these two routes can contain similar information in some "perfect" communication situations and highlighting situations where their decoding performance differs. We conclude that optimal decoding should involve some sort of fusion of association and simulation in the human brain.

<C-Text begins>
In their target article, Pickering & Garrod (P&G) propose an ambitious model of language perception and production. It is centered on three main ingredients. First, it considers the complete hierarchy of layers of language processing, from message to semantics to syntax to phonology and finally, to speech. Second, it features predictive forward models, so that temporally extended sequences, such as whole sentences and dialogues, can be processed. Third, it features dual processing routes, the "association" route and "simulation" route, so that auditory and motor knowledge can be involved simultaneously, rejecting the classic dichotomy between perception and action processes.
In this commentary, we set aside the temporal and hierarchical aspects, and focus on the domain of speech perception and production, where sequences are typically short (e.g., syllable perception and production), and processing limited to phonological decoding. Even in this more restricted field, the age-old debate between purely motor-based accounts and purely sensory-based accounts of perception and production now appears to be a false dilemma (Schwartz et al. 2012). Indeed, neurophysiological and behavioral evidence strongly suggests a dual route account of information processing in the central nervous system, with both a direct, associative route and an indirect, simulation route. The target article amply documents the evidence, we do not repeat examples here.
In our view, the debate is now shifting toward the issue of the functional role of each route and their integration. That is to say, a central question of the debate is to ask *what* is integrated and *how* integration proceeds in the human brain.
We would argue that conceptual models such as proposed in the target article would unfortunately have a difficult time bringing light to these questions. To support this argument, we consider the question of perceptual decoding of phonetic units, for which we have developed a computational framework (Moulin-Frier et al. 2012) based on Bayesian programming (Bessière et al. 2008; Colas et al. 2010; Lebeltel et al. 2004). With this framework, various models of speech perception can be simulated and quantitatively compared. One model is purely auditory, exploiting what P&G call "association." A second model is purely motor, exploiting what they call "simulation." A third one is sensory-motor, integrating the association and simulation processes.

All of these models can then be implemented and compared in various experimental configurations. Three major results emerge from such comparisons.

<<NL>(1) Under some hypotheses, with perfectly identified communication noise and no difference between motor repertoires of the speaker and the listener (i.e., when conditions for speech communication are "perfect"), motor and auditory theories are indistinguishable. Therefore, the "association" and "simulation" routes provide exactly the same information in these perfect communication conditions. The reason is that, in our learning scenario, the auditory classifier is learned by association from data obtained through a motor production process, and possesses enough mathematical power of expression.

This casts an interesting light on the question of *what* information is encoded in the association and simulation routes: Labeling a box as an "association" route, in a conceptual model, is not enough to be certain that it is different, from an information processing point of view, from another box of the model. Computational descriptions however, by virtue of rigorous mathematical notation, have to be precisely defined, and their content can be systematically assessed. This also explains why behavioral evidence has historically not been able to discriminate between motor and auditory theories of perception and production: They are sometimes simply indistinguishable. Unfortunately, we believe this difficulty was not avoided in the target article, in particular when P&G detail experimental evidence for their model (e.g., target article, sect. 3.2.1, para. 7, "these four studies support forward modeling, but they do not discriminate between prediction-by-simulation and prediction-by-association"; and sect. 3.2.3, para. 6, "all of these findings provide support for the model of prediction-by-simulation [...]. Of course, comprehenders may also perform prediction-by-association […].").

(2) In the general case where "perfect conditions" for communication are not met, mathematical comparison of the models emphasizes the respective roles of motor and auditory knowledge in various conditions of speech perception in adverse conditions. Therefore, the information provided by the "association" and "simulation" routes is more or less distinct and prominent depending on the communication conditions. In other words, this demonstrates that adverse conditions provide leverage for discriminating hypotheses about the perceptual and motor processes involved. This is convergent with

recent findings from neuroimaging and transcranial magnetic stimulation (TMS) studies (d'Ausilio et al. 2012; Meister et al. 2007; Zekveld et al. 2006), as well as computational studies (Castellini et al. 2011).

(3) In any case, sensory-motor fusion provides better perceptual performance than pure auditory or motor processes. Therefore, complementarities of information provided by the "association" and "simulation" routes could be efficiently exploited in the framework of integrative theories such as those hinted at in the discussion of the target article. It is now obvious in the field of audiovisual perception that auditory and visual cues are complementary, with a great deal of work already done on sensor fusion. In our opinion, comparable work can now be done on *how* to integrate auditory and motor processes in speech perception. In this view, the proposal by P&G that "comprehenders emphasize whichever route is likely to be more accurate" (sect. 4, para. 6) can be regarded as a first candidate model, which would have to be made mathematically precise and compared with alternative explanations, possibly driven by neuroanatomical findings (e.g., both auditory and motor processes are performed automatically in parallel and compete, or they both bring information in an ongoing fusion process, etc.).

<NL ends>

An obvious challenge, of course, is to bridge the gap between computational approaches such as ours, which are usually restricted to isolated syllable production and perception, and conceptual models as proposed in the target article, that tackle continuous flows of speech and consider semantic, syntactic and phonology layers of processing.

However, in our view, the main challenge for future studies is first to assess *what* kind of information is present in "association" and "simulation" routes, and second, to better understand *how* computational fusion models, describing the integration of these two routes, can account for experimental neurocognitive data.

<C-Text ends>

<RFT>**References[**Raphaël Laurent, Clément Moulin-Frier, Pierre Bessière, Jean-Luc Schwartz and Julien Diard][RL]

<Refs>

Bessière, P., Laugier, C. & Siegwart, R. ed. (2008) *Probabilistic reasoning and decision making in sensory-motor systems*, *volume 46 of Springer tracts in advanced robotics*. Springer-Verlag[RL]

Castellini, C., Badino, L., Metta, G., Sandini, G., Tavella, M., Grimaldi, M. & Fadiga, L. (2011) The use of phonetic motor invariants can improve automatic phoneme discrimination. *PLoS ONE* 6(9):e24055. [RL]

Colas, F., Diard, J. & Bessière, P. (2010) Common Bayesian models for common cognitive issues. *Acta Biotheoretica* 58(2–3):191–216. [RL]

D'Ausilio, A., Bufalari, I., Salmas, P. & Fadiga L. (2012) The role of the motor system in discriminating degraded speech sounds. *Cortex* 48:882–87. [RL

Lebeltel, O., Bessière, P., Diard, J. & Mazer, E. (2004) Bayesian robot programming. *Autonomous Robots* 16(1):49–79. [RL]

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D. & Iacoboni, M. (2007) The essential role of premotor cortex in speech perception. *Current Biology* 17:1692–96. [RL]

Moulin-Frier, C., Laurent, R., Bessière, P., Schwartz, J.-L. & Diard, J. (2012) Adverse conditions improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception: An exploratory Bayesian modeling study. *Language and Cognitive Processes* 27(7–8):1240–63. [RL]

Schwartz, J.-L., Basirat, A., Ménard, L. & Sato, M. (2012) The perception-for-action-control theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics* 25(5):336–54. [RL]

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M. & Schoonhoven, R. (2006) Top-down and bottom-up processes in speech comprehension. *Neuroimage* 32:1826–36. [RL]
<Refs end>