Perceptuo-motor speech units in the brain with COSMO, a Bayesian model of communication

Marie-Lou Barnaud, Julien Diard, Pierre Bessière, Jean-Luc Schwartz

Perceptuo-motor speech units in the brain with COSMO, a Bayesian model of communication

Marie-Lou Barnaud[1 2] , Julien Diard[2], Pierre Bessière[3], Jean-Luc Schwartz[1]

[1]Univ. Grenoble Alpes, Gipsa-lab, F-38000 Grenoble, France; CNRS, Gipsa-lab, F-38000 Grenoble, France

[2] Univ. Grenoble Alpes, LPNC, F-38000 Grenoble, France; CNRS, LPNC, F-38000 Grenoble, France

[3] CNRS - SORBONNE Universités - UPMC - ISIR, Paris, France

marie-lou.barnaud@gipsa-lab.grenoble-inp.fr

1. Introduction

Nowadays, three kinds of theories still discuss the nature of speech units: auditory theories (Diehl, Lotto, and Holt 2004), motor theories (Galantucci, Fowler, and Turvey 2006) and perceptuo-motor theories (Schwartz et al. 2012). Briefly, supporters of the auditory (resp. motor, perceptuo-motor) theories claim that speech units are represented in the brain by auditory (resp. motor, perceptuo-motor) information.

Several neuroimaging studies have revealed the role of the motor system during speech perception (Skipper, Devlin, and Lametti 2017). However, a recent study by Cheung et al. (2016) (noted C-2016 in the rest of the paper) questions the presumed motor nature of the activations observed in the motor cortex. By analysing high-density direct human cortical recordings, they observed a similar structure between neural patterns in the motor and auditory cortex during a perception task but different from neural patterns in the motor cortex during a production task. They conclude that the motor cortex does not contain motor but rather auditory representations in speech perception.

Intrigued by this conclusion, we assess COSMO, our Bayesian model of communication (Moulin-Frier et al. 2012), on a comparable study. Indeed, we consider that representations analyzed in the C-2016 study are not clearly identified. Specifically, the distinction between stored and computed information appears unclear. Then, by simulating an analogous perception task, we also observe a similar structure in the two cortical areas. However, this similarity disappears by using noisy stimuli, revealing the differences between representations in the auditory and motor pathways.

2. The COSMO model

The COSMO model, based on a Bayesian framework (Bessière et al. 2013), is composed of five probabilistic variables: a sound representation S, a motor gesture representation M, two objects O_L and O_S (respectively linked to S and M), and a switch C ensuring the coherence between both objects. The joint distribution $P(C\ O_L\ S\ M\ O_S)$ is decomposed into five distributions: a prior on the object $P(O_S)$, a motor system $P(M|O_S)$, a sensorimotor system $P(S|M)$, an auditory system $P(O_L|S)$ and a validation system $P(C|O_S\ O_L)$. They represent stored knowledge of the COSMO model and are employed to perform tasks.

A task, in a Bayesian model, is simulated by computing a distribution. In COSMO, we execute a speech production task with $P(M|O)$ and a perception task with $P(O|S)$. On the basis of speech perception theories, we assume that speech perception involves the fusion of an auditory decoder, operating in the auditory cortex, and which would correspond to $P(O_L|S)$, and a motor decoder, operating in the motor cortex, and which would correspond to $P(O_S|S)$. Similarly, we consider that a speech production, controlled in the motor cortex, would correspond to $P(M|O_S)$ (Laurent et al. in press).

To reproduce the C-2016 experiments, we implement a model in which O_S and O_L are four syllables /ba da pa ta/. We represent them in two dimensions, respectively refering to acoustic and articulatory features. One dimension is voicing contrasting voiced and voiceless plosives and the other is place of articulation contrasting labials and coronals. To stay with simplified hypotheses, we suppose that $P(S|O_L)$ and $P(M|O_S)$ are both Gaussian distributions, one for each object. In the voicing dimension, Gaussian distributions are far apart in $P(S|O_L)$ and close to one another in $P(M|O_S)$. It is the opposite in the place dimension. Concerning $P(S|M)$, we consider it as a sigmoid function, in accordance with the Quantal Theory of Stevens (Stevens 1989), highly non-linear in the voicing dimension and quasi-linear in the place dimension.

3. Results

We first observe that, according to the definition of our model and contrary to the claim in C-2016, production and perception tasks result in three distinct computed distributions (respectively $P(M|O_S)$ for motor cortex actvity in production, $P(O_L|S)$ for auditory cortex activity in perception and $P(O_S|S)$ for motor cortex activity in perception), each carrying different information.

Secondly, we performed perception tasks in two conditions. In the first one, named the "standard condition", we use stimuli $s^{standard}$ corresponding to prototypical syllables, comparable to those tested in the C-2016 study. This task is represented in our model by the $P(O_L|[S = s^{standard}])$ and $P(O_S|[S = s^{standard}])$ distributions. In the second condition, named the "noisy condition", we use noisy stimuli s^{noise} for which we added white noise to the standard stimuli. This task is represented in our model by the $P(O_L|[S = s^{noise}])$ and $P(O_S|[S = s^{noise}])$ distributions. In the standard condition, like in the C-2016 study, we obtained similar structures in both distributions, that is to say, the geometry of syllables in the place/voicing space is similar in auditory and motor decoding, just as C-2016 obtained, that is, similar patterns for auditory and motor cortex activity in speech perception. However, we observed in the noisy condition that both distributions are different, especially concerning the variance of the Gaussian distributions. Importantly, in COSMO, the motor decoder enables to process noisy stimuli more efficiently than the auditory decoder. It shows that auditory and motor representations would actually differ despite the similarity in activity in C-2016. Moreover, we predict that perception of noisy stimuli would actually result in different structures in the observed neural patterns.

4. Acknowledgements

Research supported by a grant from the European Research Council (FP7/2007-2013 Grant Agreement no. 339152, "Speech Unit(e)s").

5. References

Bessière, Pierre, Emmanuel Mazer, Juan Manuel Ahuactzin, and Kamel Mekhnacha (2013). *Bayesian Programming*. Boca Raton, Florida: CRC Press.

Cheung, Connie, Liberty S Hamilton, Keith Johnson, and Edward F Chang (2016). "The auditory representation of speech sounds in human motor cortex". In: *Elife* 5, e12577.

Diehl, Randy L., Andrew J. Lotto, and Lori L. Holt (2004). "Speech Perception". In: *Annual Review of Psychology* 55.1, pp. 149–179.

Galantucci, Bruno, Carol A Fowler, and Michael T Turvey (2006). "The motor theory of speech perception reviewed". In: *Psychonomic bulletin & review* 13.3, pp. 361–377.

Laurent, Raphaël, Marie-Lou Barnaud, Jean-Luc Schwartz, Pierre Bessière, and Julien Diard (in press). "A computational assessment of the role of motor information in speech perception". In: *Psychological Review*.

Moulin-Frier, Clement, Raphael Laurent, Pierre Bessière, Jean-Luc Schwartz, and Julien Diard (2012). "Adverse conditions improve distinguishability of auditory, motor, and perceptuo-motor theories of speech perception: An exploratory Bayesian modelling study". In: *Language and Cognitive Processes* 27.7-8, pp. 1240–1263.

Schwartz, Jean-Luc, Anahita Basirat, Lucie Ménard, and Marc Sato (2012). "The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception". In: *Journal of Neurolinguistics* 25.5, pp. 336–354.

Skipper, Jeremy I, Joseph T Devlin, and Daniel R Lametti (2017). "The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception". In: *Brain and language* 164, pp. 77–105.

Stevens, Kenneth N (1989). "On the quantal nature of speech". In: *Journal of phonetics* 17.1, pp. 3–45.