

# Modeling sensory preference in speech motor planning

Jean-François Patri, Pascal Perrier, Julien Diard

► **To cite this version:**

Jean-François Patri, Pascal Perrier, Julien Diard. Modeling sensory preference in speech motor planning. 11th International Seminar on Speech Production (ISSP 2017), Oct 2017, Tianjin, China. 2017. <hal-01614760>

**HAL Id: hal-01614760**

**<https://hal.archives-ouvertes.fr/hal-01614760>**

Submitted on 11 Oct 2017

**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.

# Modeling sensory preference in speech motor planning

Jean-François Patri<sup>\*1,2</sup>, Pascal Perrier<sup>†1</sup>, and Julien Diard<sup>‡2</sup>

<sup>1</sup>Univ. Grenoble Alpes, CNRS, Gipsa-lab, UMR 5216 F-38000 Grenoble, France

<sup>2</sup>Univ. Grenoble Alpes, CNRS, LPNC UMR 5105, F-38000 Grenoble, France

Speech is a stream of specific sounds performed by gestures of articulators of the vocal tract. The sensory correlates of speech production are therefore both auditory (concerning sounds) and somatosensory (concerning the position and configuration of articulators of the vocal tract). Since sounds are a consequence of speech gestures, these two sensory correlates appear to be redundant in unperturbed conditions. This raises questions about their functional involvement in the monitoring of speech production: is only one useful, and if so, which one? Are they instead both useful, and if so, are they equivalent or complementary?

Experimental studies of compensations for auditory and somatosensory perturbations indicate that both types of sensory information are taken into account during speech production [6, 2]. In addition, individual sensory preferences in speech production have been observed: subjects who compensate less for somatosensory perturbations compensate more for auditory perturbations, and vice versa [3]. Our goal is to understand how sensory preferences can operate during speech production and influence it, by using our recently designed Bayesian model of speech motor planning. To our knowledge, models of speech motor control have generally not addressed this issue since they did not systematically evaluate the consequences of variations in the weight of each modality in the specification of the motor goals.

In this work, we present extensions of our original Bayesian model of speech motor planning [4], in which speech units (phonemes in our case) are characterized both in auditory and somatosensory terms. Figure 1 presents a graphical representation of the model. Variable  $\Phi$  corresponds to phonemes which are characterized in terms of auditory and somatosensory variables  $A_\Phi$  and  $S_\Phi$ . The dependencies relating these variables are probability distributions  $P(A_\Phi | \Phi)$  and  $P(S_\Phi | \Phi)$  which we identify to multivariate normal distributions parameterized by means and covariances  $(\mu_A^\phi, \Sigma_A^\phi)$  and  $(\mu_S^\phi, \Sigma_S^\phi)$  respectively. Variables  $A_M$  and  $S_M$  represent the predicted auditory and somatosensory consequences attributed to motor commands  $M$ . Variables  $C_A$  and  $C_S$  act as two sensory-matching-constraint indicators, displaying value 1 when values of the corresponding sensory variables ( $A_\Phi$  and  $A_M$  or  $S_\Phi$  and  $S_M$ ) match and 0 otherwise. The planning of motor commands  $M$  for the production of an intended phoneme  $\phi$  is identified to the outcome of the inference question  $P(M | [\Phi = \phi] [C_A = 1] [C_S = 1])$ , imposing the identity  $\phi$  of the phoneme and that both sensory-matching constraints are satisfied (in other words  $[C_A = 1]$  and  $[C_S = 1]$ ). This results in a planning process that takes into account both the auditory and the somatosensory characterization of phonemes.

We show that sensory preferences can be modeled in two ways in our framework. In the first variant, sensory preferences are attributed to the relative precision of sensory regions characterizing speech motor goals,  $P(A_\Phi | \Phi)$  and  $P(S_\Phi | \Phi)$ . Precision correspond in our framework to the inverse of covariance parameters  $\Sigma_A^\phi$  and  $\Sigma_S^\phi$ . This approach is inspired from classical models of multisensory fusion for perception [1, 5]. Under this approach, precision of sensory regions correspond to their tolerance to perturbations: the smaller the region, the higher the precision and the lower the tolerance to perturbations. In other words, subjects who compensate more to auditory than somatosensory perturbations would have auditory target regions smaller than their somatosensory target regions. However, since auditory and somatosensory consequences of speech gestures are highly correlated, why would these motor goal regions differ so considerably?

In the second variant of our model, sensory preferences are the consequence of the relative precision by which the predicted sensory consequences of motor commands are compared to the sensory characterizations of motor goals. This comparison arises in the dependencies relating the sensory-matching-constraint variables  $C_A$  and  $C_S$  and their corresponding sensory variables ( $A_\Phi$  and  $A_M$  or  $S_\Phi$  and  $S_M$ ) which we define as

$$P([C_A = 1] | [A_\Phi = a_\phi] [A_M = a_m]) \propto e^{-\frac{\|a_\phi - a_m\|^2}{2\sigma_a^2}}, \quad (1)$$

$$P([C_S = 1] | [S_\Phi = s_\phi] [S_M = s_m]) \propto e^{-\frac{\|s_\phi - s_m\|^2}{2\sigma_s^2}}, \quad (2)$$

---

\*Jean-Francois.Patri@gipsa-lab.fr

†Pascal.Perrier@gipsa-lab.fr

‡julien.diard@univ-grenoble-alpes.fr

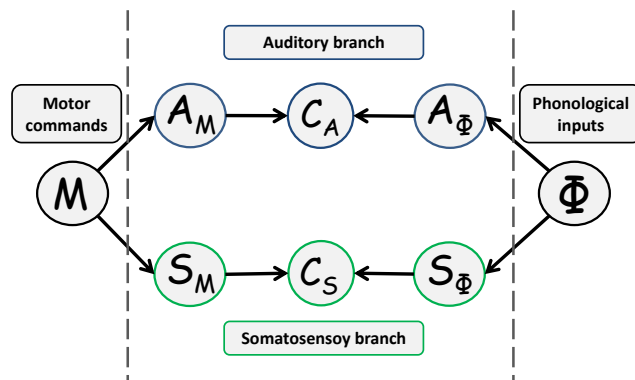


Figure 1: Graphical representation of our Bayesian model for speech motor planning; nodes represent variables and arrows dependency relations. Variable  $\Phi$  corresponds to phonemes which are characterized in terms of auditory and somatosensory variables  $A_\Phi$  and  $S_\Phi$ . Variables  $A_M$  and  $S_M$  represent the predicted auditory and somatosensory consequences attributed to motor commands  $M$ . Variables  $C_A$  and  $C_S$  implement two sensory-matching constraints and act as binary switches activating or not the corresponding sensory pathway.

where operator  $\|\cdot\|$  denotes the Euclidean norm. In other words, the probability of variable  $C_A$  (respectively  $C_S$ ) taking value 1 (indicating sensory-matching) is maximum when values of variables  $A_\Phi$  and  $A_M$  (respectively  $S_\Phi$  and  $S_M$ ) are the same and decreases exponentially as they get separated. Parameters  $\sigma_a$  and  $\sigma_s$  modulate the exponential decrease with respect to the distance between sensory values: the smaller  $\sigma_a$  (respectively  $\sigma_s$ ), the stronger the decrease in probability for a given distance between values of  $A_\Phi$  and  $A_M$  (respectively  $S_\Phi$  and  $S_M$ ).

We demonstrate that this second implementation of sensory preferences is formally equivalent to an increase of parameters of the covariance matrices,  $\Sigma_A^\phi$  and  $\Sigma_S^\phi$ , that specify the sensory characterizations of phonemes,  $P(A_\Phi | \Phi)$  and  $P(S_\Phi | \Phi)$ . This relates and reconciles the two approaches and suggests an alternative and original interpretation of sensory preferences, where target precision are not modulated, but instead, tolerance to error in auditory or somatosensory spaces are different. We further illustrate our two approaches by exploring the outcome of adaptation to auditory and somatosensory perturbations in the context of a 2-dimensional bio-mechanical model of the tongue developed in our lab.

## References

- [1] Marc O Ernst and Martin S Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002.
- [2] John F Houde and Michael I Jordan. Sensorimotor Adaptation in Speech Production. *Science (New York, N. Y.)*, 1(5354):1213–1216, 1998.
- [3] Daniel R Lametti, Sazzad M Nasir, and David J Ostry. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(27):9351–9358, 2012.
- [4] Jean-François Patri, Julien Diard, and Pascal Perrier. Optimal speech motor control and token-to-token variability: a Bayesian modeling approach. *Biological Cybernetics*, 109(6):611–626, 2015.
- [5] Joseph S Perkell, Harlan Lane, Satrajit Ghosh, Melanie L Matthies, and Haskins Laboratories. Mechanisms of Vowel Production : Auditory Goals and Speaker Acuity. In *Proceedings of the Eighth International Seminar on speech production, Strasbourg, France*, pages 29–32, 2008.
- [6] S Tremblay, D M Shiller, and D J Ostry. Somatosensory basis of speech production. *Nature*, 423(6942):866–869, 2003.