Body Representations for Robot Ego-Noise Modelling and Prediction. Towards the Development of a Sense of Agency in Artificial Agents

Guido Schillaci[1], Claas N. Ritter[1], Verena V. Hafner[1], Bruno Lara[2]

[1]Adaptive Systems Group, Department of Computer Science, Humboldt-Universität zu Berlin, Berlin, Germany
[2]Cognitive Robotics Group, Center for Science Research, Universidad Autonoma del Estado de Morelos, Cuernavaca, Mexico
Corresponding author: guido.schillaci@informatik.hu-berlin.de

Abstract

We present an implementation of a biologically inspired model for learning multimodal body representations in artificial agents in the context of learning and predicting robot ego-noise. We demonstrate the predictive capabilities of the proposed model in two experiments: a simple ego-noise classification task, where we also show the capabilities of the model to produce predictions in absence of input modalities; an ego-noise suppression experiment, where we show the effects in the ego-noise suppression performance of coherent and incoherent proprioceptive and motor information passed as inputs to the predictive process implemented by a forward model. In line with what proposed by several behavioural and neuroscience studies, our experiments show that ego-noise attenuation is more pronounced when the robot is the owner of the action. When this is not the case, sensory attenuation is worse, as the incongruence of the proprioceptive and motor information with the perceived ego-noise generates bigger prediction errors, which may constitute an element of surprise for the agent and allow it to distinguish between self-generated actions and those generated by other individuals. We argue that these phenomena can represent cues for a sense of agency in artificial agents.

Introduction

Empirical evidence from cognitive science and neuroscience suggests that we, as humans, maintain an internal representation of our body, or a model of our motor system, and that such an internal model would be involved in processes of simulation of sensorimotor activity. These processes would affect the way we experience the interaction with the environment and would be fundamental for the implementation of basic cognitive skills. For example, simulation processes are thought to have a role in the way we differently perceive self-generated actions or actions performed by other subjects. One of the proposals that explains this phenomenon (Blakemore et al., 2000a,b) says that when we perform a motor action, an efferent copy of the motor commands that our brain sends to our muscles would be used in a predictive process that anticipates the sensory outcomes of the movement. Such predictions would be then compared to the actual sensory consequences and, if the two correspond, the perceived sensory consequences are attenuated. This would enable a differentiation between self-generated sensory events and those externally generated that are not mapped to any internally generated efferent copy of the motor commands (Blakemore et al., 2000a). The existence of such a self-monitoring mechanism would explain, for example, why tickling sensations cannot be self-produced (Blakemore et al., 2000b), why people are better at recognising themselves than others when watching movies of only point-light walkers (Casile and Giese, 2006), why people are more accurate in predicting the landing point of a thrown dart from a video screen when they observe their own throwing action than when observing another person's throwing action (Knoblich and Flach, 2001), or why people perceive the loudness of sounds as less intensive when they are self-generated, than when they are generated by other persons or by a software (Weiss et al., 2011). In this latter study on selective attenuation of self-generated sounds, the authors proposed that the experience of perceiving actions as self-generated would be caused by the anticipation and, thus, the attenuation of the sensory consequences of such motor commands, which would be related to "the privileged access to internally generated efferent information during one's own action" (Weiss et al., 2011). The sense of agency, that is the pre-reflective experience that *we* are the owner of an action we are executing, is thus proposed to be dependent on the degree of congruence vs. incongruence between predicted and actual sensory consequences of our bodily actions.

In the investigation on sensorimotor simulation processes in the human brain, internal forward and inverse models have been proposed (Wolpert et al., 2001). A forward model (illustrated in Figure 1) - or predictor, as firstly proposed in the control literature as a means to overcome problems such as the delay of feedback in control strategies (Jordan and Rumelhart, 1992) - incorporates knowledge about sensory outcomes of self-generated actions. Inverse models (illustrated in Figure 2) - or controllers, as they were initially proposed for implementing inverse kinematics processes for controlling robotic manipulators - perform the opposite transformation providing a system with the necessary motor command to go from an initial sensory situation to

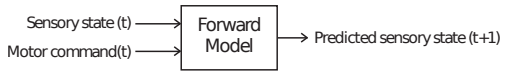

Figure 1: An illustration of the forward model (predictor).

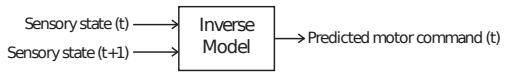

Figure 2: An illustration of the inverse model (controller).

a desired one. Such models encode the dynamics of the motor system and can provide artificial agents with multimodal representations, as they fuse together sensory and motor information (Wilson and Knoblich, 2005), and with the capability to predict sensorimotor activities based on previous experience. Studies such as the ones reported above shed light on the importance that predicting sensory consequences of self-generated actions have for basic motor tasks and cognitive skills. Equipping artificial agents with similar computational processes has been shown to be a promising approach in the development of different skills, such as navigation (Möller and Schenck, 2008; Escobar et al., 2012), perception of the functional role of objects (Kaiser, 2014), action selection and tool-use (Schillaci et al., 2012) and sense of agency (Pitti et al., 2009).

The work presented here adopts a biologically inspired framework for internal body representations (Schillaci et al., 2014) that can enable a robot with the capability to perform simulations of sensorimotor activities based on previous experience. Inspired by human development, the learning of this body representation is intertwined with the interaction experience of the robot with the external environment.In particular, we frame this work into the context of one of the biggest - and most unexplored - challenges of robot audition, the artificial capability of listening, that is the presence of *ego-noise*, or the noise that the robot generates while moving around. Being able to estimate self-induced changes in the auditory signal is not only crucial for attenuating the noise, and thus for enhancing the auditory signal for further processing such as speech recognition, but also for distinguishing ego-noise from other sounds in natural acoustic environments, which is a prerequisite for efficient and intuitive interaction with other people and with the surrounding.

We demonstrate the predictive capabilities of the model in the auditory domain in two tasks. Firstly, we introduce the framework in a simple classification task, where the robot has to recognize a behaviour that it executes based on the comparison of the produced ego-noise to internal simulations of ego-noise produced by *intended* actions. We show also how our model can deal with the situation when input information are missing, for example by simulating a damage in the system, resulting in the model still being able to classify, although with poorer performance. Secondly, we show how the proposed framework, and the predictive capabilities that it provides, could serve as a basis for the development of a sense of agency in artificial agents. In particular, we report an experiment on ego-noise attenuation based on sensorimotor predictions, where the quality of the attenuation is dependent on the degree of congruence vs. incongruence between predicted and actual sensory consequences of self-generated actions. In line with the behavioural studies reported above, we show that prediction errors generated by internal sensorimotor simulations are smaller when the proprioceptive information is coherent with the events that are perceived from the external environment. Simply put, we show that sensory attenuation is more pronounced when the robot is the owner of the action, and we argue that this could serve as a cue for self-agency in artificial agents.

In the rest of the paper we firstly introduce the framework presented in (Schillaci et al., 2014) and extend it. Therefore, we illustrate and discuss the experiments mentioned above. Finally, we draw the conclusions and the outlines of future work.

An Internal Body Representation for a Humanoid Robot

Evidences from behavioural sciences and neuroscience suggest that motor and brain development are strongly intertwined with the experiential process of exploration, where internal body representations would be formed and maintained over time (Cang and Feldheim, 2013). Kaas (1997) reported the existence of topographic maps in the visual, auditory, olfactory and somatosensory systems, as well as in parts of the motor brain areas. Researchers proposed that such maps would self-organise throughout the brain development and along the sensorimotor experience of the individual with the external environment. They would function as projections of sensory receptors and of effector systems, and are arranged in a way that adjacent regions process spatially close sensory parts of the body. Many studies supported the existence of an integrated representation of visual, somatosensory, and auditory peripersonal space in human and non-human primates (see for example Holmes and Spence (2004)), suggesting that the brain maintains integrated multimodal representations, which are essential for sensorimotor control (Maravita and Iriki, 2004).

During the last couple of decades, interest in the possibility to develop models inspired by the mechanisms of human body representations has been growing also in the robotics community. In robot audition, for example, Ince and colleagues investigated methods for learning, predicting and suppressing robot ego-noise (Ince et al., 2009). The authors built up an internal body representation of a humanoid robot consisting in motor sequences mapped to the recorded motor noises and their spectra. This resulted in a large noise template database that was then used for ego-noise prediction and subtraction.

Here, we report an implementation of a biologically inspired model for body representations that can encode experience gathered through sensorimotor learning and that can generate predictions of auditory and motor states. In particular, we propose an internal models framework consisting of connected neural networks that simulate distinct sensorimotor brain areas. The internal model encodes sensory and motor modalities as topographic maps that self-organise throughout the interaction of the robotic agent with the external environment. Moreover, a parallel intermodal mapping is performed: sensory and motor maps are connected through Hebbian links that are strengthened when an occurrence of multi-modal activity is observed.

The model architecture is inspired by the Epigenetic Robotics Architecture (Morse et al., 2010), where a structured association of multiple Self-Organising Maps (SOMs) (Kohonen, 1982) is adopted for mapping different sensorimotor modalities in a humanoid robot, and it is based on similar works we previously published (Kajić et al., 2014; Schillaci et al., 2014). Self-organising maps have the advantage of producing low-dimensional and discretised representations of the input space of the training samples.

In the proposed model, multiple SOMs, each representing a sensory or motor modality, are associated through unidirectional Hebbian links: each node of the input map is connected to a each node of the output map, where the connection is characterised by a weight. The weight is updated according to a positive Hebbian rule that simulates synaptic plasticity of the brain: the connection between a pre-synaptic neuron (a node in the input map) and a post-synaptic neuron (a node in the output map) increases if the two neurons are simultaneously activated. Learning of the internal model consists in updating the SOMs and the Hebbian connections with sensory and motor data gathered through an exploration behaviour executed by the robot. During the execution of the robot movements, sensory and motor data are provided as training inputs to the corresponding maps in an online fashion. A SOM is constructed as a grid of neurons, where each neuron is represented as an n-dimensional weight vector $\mathbf{w_i}$ (Kajić et al., 2014; Kohonen, 1982). The number of dimensions of a weight vector corresponds to the dimensionality of the input data. Weights in the network are initially set to random values and then adjusted iteratively by presenting the input vector $\mathbf{x_p}$. In each iteration, the winning neuron i is selected as a neuron whose weights are closest to the input vector in terms of the Euclidean distance. After selecting a winning neuron, the weights of all neurons are adjusted:

$$\mathbf{\Delta w_j} = \eta(t)h(i,j,t)(\mathbf{w_j} - \mathbf{x_p}) \tag{1}$$

The parameter $\eta(t)$ is a learning rate which defines the speed of change. The function $h(i,j,t)$ is a Gaussian neighborhood function defined over the grid of neurons as:

$$h(i,j,t) = e^{\left(\frac{\mathbf{w_i^2} - \mathbf{w_j^2}}{2\pi\sigma(t)^2}\right)} \tag{2}$$

The learning rate $\eta(t)$ and the spread of the Gaussian function $\sigma(t)$ are held constant for a certain time interval, and are annealed exponentially afterwards.[1] The function is centered around the winning neuron i and its values are computed for all neurons j in the grid. The spread of the function determines the extent to which neighbouring weights of a winning neuron are going to be affected in the current iteration. The topology of the network is preserved by pulling together neurons towards the winning node.

After every update of the SOMs, the Hebbian links connecting each pair of maps are updated as well. The Hebbian update corresponds to the following steps. For mapping an input map (e.g. the motor map) to an output map (e.g. the auditory map):

- select the pre-synaptic neuron (winner node) as the closest node i in the input map to the current input pattern $\mathbf{x}$ (e.g. the joint rotation);
- select the post-synaptic neuron (winner node) as the closest node j in the output map to the current output pattern $\mathbf{y}$ (e.g. the robot ego-noise);
- strengthen the connection w_{ij} between the pre and post-synaptic neurons according to the modified positive Hebbian rule:

$$\Delta w_{ij} = \lambda A_i(\mathbf{x}) A_j(\mathbf{y}) \tag{3}$$

where $A_i(\mathbf{x})$ is the activation function of the neuron i over the Euclidean distance between the neural weights and the data pattern x, λ is a scaling factor for slowing down the growth of the weights (in the experiments presented here, it is initialised to 0.1). The activation function of a neuron, $A(\mathbf{d})$, is computed as:

$$A(\mathbf{d}) = \frac{1}{1 + 2 * \tanh(\mathbf{d})} \tag{4}$$

where $\mathbf{d}$ is the normalised Euclidean distance between the position of the node and the input pattern.

After the update, a normalisation is performed on all the links from the input map converging to a node in the output map, for each node in the output map, as described by Miikkulainen (1990). Such a normalisation implements a forgetting process, since it strengthens the updated link and it weakens all the other connections. The same process is performed on the unidirectional links connecting each pair of maps in the model in both directions.

The trained model can be used for performing sensory and motor predictions. Predictive processes can be activated by

[1] In the experiments presented in the the following section, we set $\eta = 0.9$ and $\sigma = 0.7$.

querying the model with partial or full sensorimotor information. For example, we can infer the ego-noise produced by the execution of a specific motor command (forward prediction) from the model depicted in Figure 3 by querying the model with an input to the proprioceptive map, consisting of the joints configuration of the robot, and an input to the motor map, consisting of the joints rotations, which are therefore propagated to the auditory map. In fact, a predictive system based on propagation of signals between maps has been implemented. The propagation of signals works as follows. Given a sensory or motor input:

- Find the winner node w and its k neighbors (k set to 5, in the experiments presented here) in the corresponding map, as the closest node to the input, and calculate its activation using the activation function described in (4);
- Propagate the activation of the nodes in the winners list of the input map to all the nodes in the output map connected to it. The propagated value to each node in the output map is equal to the activation of the selected node in the input map multiplied by the weight of the Hebbian link connecting the selected node in the input map to the corresponding node in the output map; multiple propagations to the same node in the output map are summed up;
- Compute the prediction in the output modality as the weighted average of the positions of the nodes in the output map, each weighted by the incoming propagation.

If an observation of the output modality is available, a *prediction error* can be computed as the distance between the predicted outcome and the observation.

Moreover, multiple propagations can be executed from different input modalities to the same output modality, as illustrated in Figure 3. From each input modality, signals can be spread out to the desired output modality as described above. Thus, incoming propagations onto the output map can be summed up and a prediction can be computed as the weighted average of the nodes' positions multiplied with their activations.

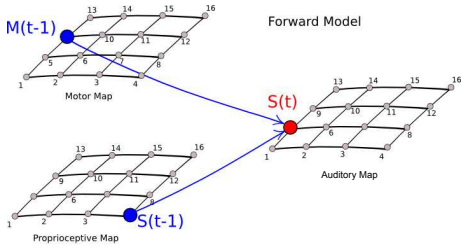

Figure 3: An example of a forward model consisting of three maps.

Figure 3 illustrates a forward model (as described in Figure 1) implemented using the proposed architecture composed of three SOMs: a proprioceptive map, encoding the initial joint configuration of the robot, a motor map, i.e. encoding the rotation applied to the joints from the initial positions, and an auditory map, encoding the noise produced by the movements. An inverse model can be implemented with two sets of directional Hebbian links: the first starting from the proprioceptive map and ending to the motor map, and the second starting from the acoustic map and ending to the motor map.

Ego-noise representation

We represent the ego-noise produced by the robot movements using Mel-frequency cepstral coefficients (MFCCs), which are features derived from a type of cepstral representation of the auditory signal commonly used in speech recognition (Sahidullah and Saha, 2012).

In this work, MFCC features are derived performing the following steps:

- Calculate the Fourier transform of an audio chunk extracted from the input signal. In the experiments reported here, we used a single channel audio signal, recorded from the robot with sampling rate of 48 kHz. Audio chunks of 40 ms are extracted from the signal using a rectangular window. Chunks are extracted every 20 ms (that is, with a 50% overlapping between subsequent chunks). FFT size is 2048 samples. 32 triangular overlapping filters are used in the Mel filterbank, with a mel filter width of 200. The frequency range of the filterbank goes from 0 to 16 kHz;
- Apply the Mel filterbank to the power of the spectrum and sum the energy in each filter;
- Calculate the Discrete Cosine Transform of the logarithm of the filterbank energies;
- Keep the first 26 or 32 coefficients of the DCT as MFCC features.

For implementing the MFCC feature extraction process, we adopted and extended an existing open source and cross-platform digital signal processing library, named Aquila DSP (http://aquila-dsp.org/).

Before being processed, input data streams are aligned in time, to ensure that the auditory stream matches the actions executed by the robot. We use the NAOqi and experimental NAOqi-Modularity frameworks provided by Aldebaran Robotics, which allow us to combine asynchronous data collection and data processing using filter chains in the humanoid robot Nao.

Experiments

We report here two experiments. Firstly, we present a simple classification experiment with the aim of demonstrating the learning and predictive capabilities that the proposed model can provide to artificial agents. In particular, we adopt the proposed framework for allowing a humanoid robot to learn the ego-noise that it is producing when performing a motor behaviour consisting of periodical horizontal head rotations (see Figure 4). Thus, we implement a classification

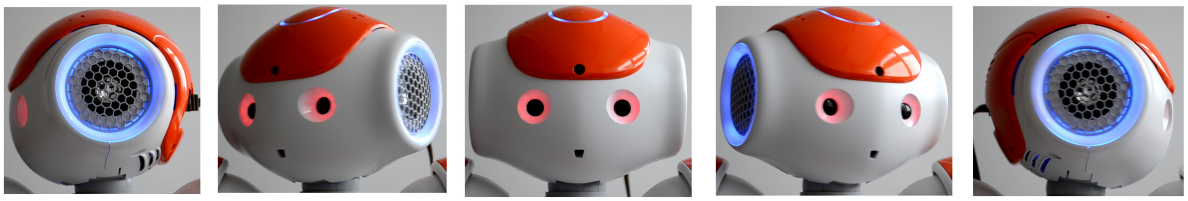

Figure 4: The robot behaviour executed during the recordings consisted in periodical rotations of the head on the yaw axis.

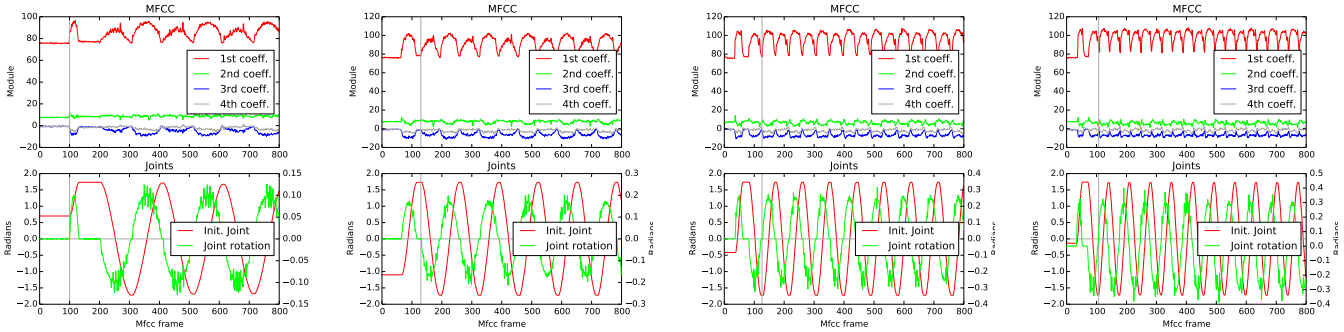

Figure 5: Example of trajectories of synchronised audio-motor data from the four velocity profiles. The upper plot shows the trajectories of the first 4-MFCC coefficients extracted from the single channel auditory signal recorded while executing the head rotations. The plot in the bottom shows the head yaw joint position (red line) and the head yaw rotation over $40ms$ (green line). The columns represent the different velocity profiles. From left to right: *slow*, *medium*, *fast* and *very fast*.

experiment where the robot has to classify a behaviour it is executing in terms of velocity profile, by comparing the produced ego-noise to simulations of the ego-noise produced by imaginary executions of all the behaviours in the repertoire. In addition, we show how the model can deal with the situation when input information are missing, for example due to a damage in the system, resulting in the model still being able to classify, although with poorer performance.

Secondly, we describe an experiment on ego-noise attenuation with the aim of showing that the computational processes implemented by our framework resemble those proposed by the behavioural studies mentioned in the introduction of this work, which would explain the mechanisms behind the sense of agency (Weiss et al., 2011; Blakemore et al., 2000a). In particular, we report an experiment on ego-noise attenuation based on sensorimotor predictions, where the quality of the attenuation is dependent on the degree of congruence vs. incongruence between predicted and actual sensory consequences of self-generated actions. In line with the behavioural studies reported above, we show that prediction errors generated by internal sensorimotor simulations are smaller when the proprioceptive and motor information are coherent with the events that are perceived from the external environment. We reported similar results in a different robotic experiment in the context of visuo-motor coordination (Schillaci et al., 2013).

Ego-noise classification

In the first experiment, we trained four different models with sensorimotor data gathered while executing a robot behaviour consisting of periodical horizontal head rotations with four different velocity profiles. We implemented the four velocity profiles using the original Aldebaran NAOqi controller with gradually increasing velocity thresholds, here named as *slow*, *medium*, *fast* and *very fast*. Figure 5 shows sample trajectories of aligned auditory and motor training data for each of the four velocity profiles. Training of the models has been tested online and runs in real-time on an Aldebaran Nao v.5 robot. However, the classification results reported here are taken from models trained and tested offline. Sensorimotor data was gathered from the robot executing for ca. 200 seconds each of the four velocity profiles, resulting in 9449 training samples for the *slow* velocity profile, 9449 for the *medium*, 9459 for the *fast* and 9459 for the *very fast*. Each training sample consisted of the following sensorimotor information:

- S(t): MFCC features extracted from a single audio chunk (see Section "Ego-noise representation" for more details);
- S(t-1): initial position of the head yaw joint, that is the closest position in time to the first audio sample of the MFCC chunk;
- M(t-1): rotation of the head yaw joint over 40 ms, from S(t-1).

Four internal models have been trained with the different datasets (slow, medium, fast and very fast velocity profiles). Each internal model consisted of three maps (see Figure 3): a proprioceptive map, encoding a mono-dimensional feature space representing the initial head yaw joint position, that is $S(t-1)$; a motor map, encoding a mono-dimensional feature space representing the head yaw joint rotation, that is the motor command $M(t-1)$; an auditory map, encoding a 26-dimensional MFCC feature space representing the robot ego-noise. Each internal model encoded both the inverse and the forward models, as these are implemented by the Hebbian tables containing the proper directional links, as explained in the previous section. Each SOM consisted of a 10x10 lattice of nodes, whose weights are randomly initialised and sampled from a Gaussian distribution $\mathcal{N}(0,1)$. The weights of the Hebbian links connecting each pair of SOMs were all initialised to 0.

The classification task consisted in feeding the four internal models (slow, medium, fast and very fast) with test data samples gathered from the different datasets which stored sensorimotor data produced with each of the four velocity profiles, and in comparing the predicted auditory outcome with the actual one. Auditory chunks are classified as the velocity profile belonging to the forward model that produced the smallest prediction error (calculated as the Euclidean distance between the predicted and the observed MFCCs). Figure 6 illustrates the classification process using internal simulations.

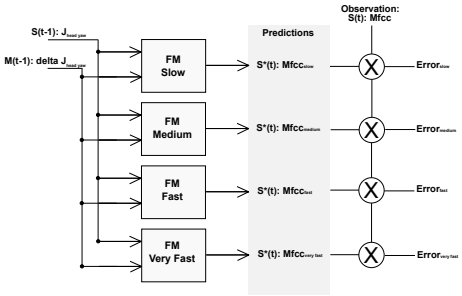

Figure 6: Diagram of the classification process.

Classification performance was measured for each trained model and on 5 different runs (thus, different test datasets). Table 1 shows the confusion matrix for the best run, when using only forward predictions with full input information.

		Classified as				
		Slow	Medium	Fast	Very fast	# samples
Executed velocity	Slow	**94,00%**	5,50%	0,50%	0,00%	200
	Medium	6,00%	**89,50%**	4,00%	0,50%	200
	Fast	0,50%	14,00%	**85,5%**	0,00%	200
	Very Fast	4,00%	4,50%	7,50%	**84,00%**	200

Table 1: Confusion matrix showing the performance of the classification using only forward predictions.

Thus, we simulated a damage in the system, which was implemented as a lack of proprioceptive and motor information, during the predictive process. Internal simulations with partial inputs - in this case, only the auditory modality - were performed. The first step consisted in estimating a prediction of the motor command needed to generate the auditory outcome specified as input to the model, using an inverse prediction. The predicted motor command is thus fed into the corresponding forward model, which anticipates the sensory outcome of the intended action. Table 2 shows the confusion matrix of the best of 5 classification runs, where we executed full internal simulations using only partial information as input. As expected, predictions estimated with missing proprioceptive inputs produced a degradation of the classification performace. However, the system is still able to classify correctly with at least 50% success.

		Classified as				
		Slow	Medium	Fast	Very fast	# samples
Executed velocity	Slow	**85,5%**	11,50%	2,50%	0,50%	200
	Medium	7,50%	**77,50%**	13,00%	2,00%	200
	Fast	1,00%	16,00%	**80,0%**	3,00%	200
	Very Fast	1,50%	7,50%	37,00%	**54,00%**	200

Table 2: Confusion matrix showing the performance of the classification using both the inverse and forward predictions with missing input data (proprioceptive joints information).

Ego-noise attenuation as a cue for sense of agency

We performed a second experiment on ego-noise attenuation based on ego-noise predictions. In the experiment, we simulated that the robot is listening to a ego-noise signal (previously recorded from the robot itself) and, in the meanwhile, performing a motor behaviour. Along these movements, a forward model - trained with a periodical head rotation behaviour with slow velocity profile, as in the previous experiment - was used in executing sensorimotor simulations aimed at predicting the robot ego-noise generated by the current motor behaviour of the robot. We tested three conditions. In the first one, we simulated that the robot is executing a motor behaviour that is coherent with the observed ego-noise. In a second one, we simulated that the robot is not moving, thus holding the head in an initial position (applying a null motor command). In a third condition, we simulated that the robot is performing a periodical head rotation that is not aligned in time with the observed ego-noise. In each of the three conditions, we predicted the auditory outcomes of the movements by feeding the forward model with the joints information corresponding to the current motor behaviour. Thus, we subtracted from the original auditory information the one of the estimated noise.

Ego-noise suppression is performed in the log-filterbank energies domain. An inverse DCT (Discrete Cosine Transform) is applied to the 32-MFCC feature vectors representing the predicted and actual ego-noise chunks, producing

two 32-D vectors (log-filterbank energies). Therefore, the vector representing the predicted ego-noise is subtracted from the one representing the actual ego-noise. In the event that the subtraction result in a dimension is negative, spectral flooring is applied, that is the attenuated signal is computed as the original one multiplied with a factor of 0.1.

Figure 7 qualitatively illustrates the results of the ego-noise attenuation. As evident from the plots, ego-noise attenuation is more pronounced when the input data fed to the forward model is coherent with the auditory output (left graphs in the Figure - dark blue colour corresponds to total suppression of the ego-noise). The quality of the attenuation is worse, when there is incongruence between predicted and actual sensory consequences of self-generated actions, as in the case of the second and third condition. In particular, the second behaviour (head holding an initial position) generates a constant ego-noise prediction. The difference between the original and predicted ego-noise (bottom row, central column) is thus higher than in the case when the motor behaviour matches the observed ego-noise. Same effect is observed in the third condition, where the motor behaviour does not match the observed ego-noise.

In line with the studies reported in the introduction of this study, our experiment shows that prediction errors generated by sensorimotor simulations are smaller when the proprioceptive and motor information are coherent with the perceived ego-noise. Simply put, sensory attenuation is more pronounced when the robot is the owner of the action, as it has "a privileged access to internally generated efferent information during its own action" (Weiss et al., 2011), as simulated in the first condition of this experiment. The second and third condition simulated the situation where the robot is listening to another artificial agent performing a periodical horizontal head rotation behaviour, that *sounds exactly* as it would have been produced by the robot itself. However, the fact that the observed proprioceptive and motor information were incoherent with the observations of the ego-noise did constitute an element of surprise, as the forward model fed with such input data produced worse ego-noise prediction than in the first condition of the experiment - as evident from the bigger prediction errors illustrated in Figure 7, bottom plots of the second and third columns.

Conclusion

We presented an implementation of a biologically inspired model for coding internal body representations that can generate predictions of auditory and motor experiences. The predictive capabilities provided by the models are tested in two experiments: a simple ego-noise classification task, where we also showed the capabilities of the model to produce predictions even in the absence of input modalities; an ego-noise suppression experiment, where we showed the effects in the ego-noise prediction, and thus suppression, performance of the input data to the forward model, when they are coherent or incoherent with the auditory observations. In line with the behavioural studies reported in the introduction of this study, our experiment shows that prediction errors generated by sensorimotor simulations are smaller when the proprioceptive and motor information are coherent with the perceived ego-noise. Simply put, sensory attenuation is more pronounced when the robot is the owner of the action. When this is not the case, sensory attenuation is worse, as the incongruence of the proprioceptive and motor information with the perceived ego-noise generates bigger prediction errors, which may constitute an element of surprise for the agent and allow it to distinguish between self-generated actions and those generated by other individuals. Therefore, we argue that equipping artificial agents with internal body representations and with the capability to perform sensorimotor predictions based on previous experience can represent a promising research direction towards the development of a sense of agency in artificial systems.

Acknowledgments. The research leading to these results has partially received funding from the European Unions Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 609465 (EARS (Embodied Audition for RobotS) Project).

References

Blakemore, S.-J., Smith, J., Steek, R., Johnstone, E. C., and Frith, C. D. (2000a). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychological Medicine*, 30:1131–1139.

Blakemore, S. J., Wolpert, D., and Frith, C. (2000b). Why can't you tickle yourself? *NeuroReport*, 11(11):11–16.

Cang, J. and Feldheim, D. A. (2013). Developmental mechanisms of topographic map formation and alignment. *Annual Review of Neuroscience*, 36(1):51–77. PMID: 23642132.

Casile, A. and Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Current Biology*, 16(1):69 – 74.

Escobar, E., Hermosillo, J., and Lara, B. (2012). Self body mapping in mobile robots using vision and forward models. In *Electronics, Robotics and Automotive Mechanics Conference (CERMA)*, pages 72–77.

Holmes, N. and Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cognitive Processing*, 5(2):94–105.

Ince, G., Nakadai, K., Rodemann, T., Hasegawa, Y., Tsujino, H., and Imura, J. (2009). Ego noise suppression of a robot using template subtraction. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2009)*, pages 199–204.

Jordan, M. I. and Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science*, 16:307–354.

Kaas, J. H. (1997). Topographic maps are fundamental to sensory processing. *Brain Research Bulletin*, 44(2):107 – 112.

Kaiser, A. (2014). *Internal visuomotor models for cognitive simulation processes*. PhD thesis, Bielefeld University.

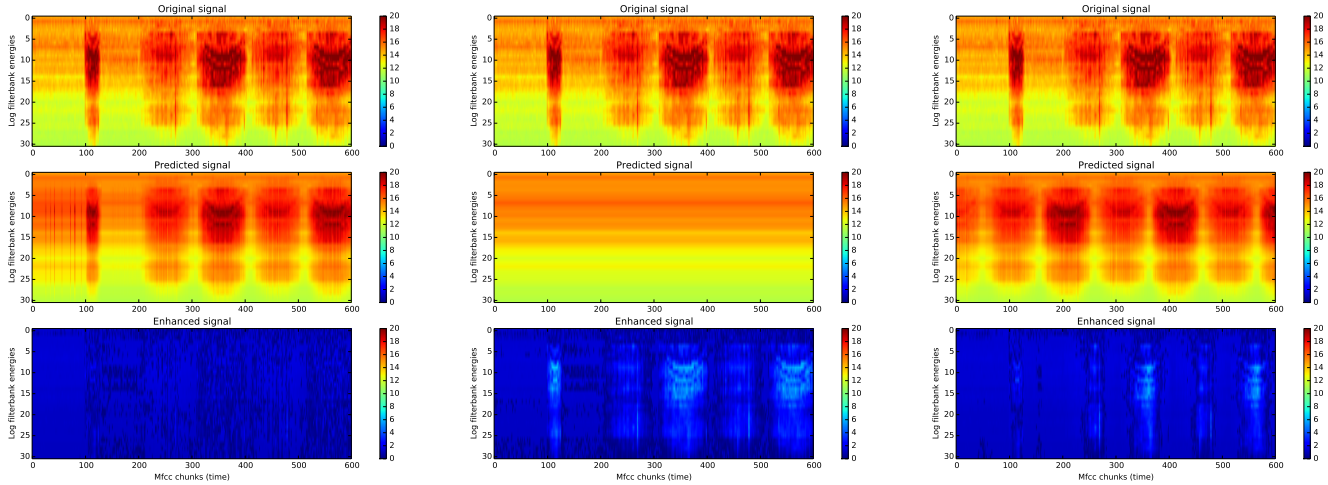

Figure 7: Ego-noise prediction and suppression tests. In the left graphs, input joint states and motor commands are coherent with the auditory outcome of the movements. In middle and right graphs, these inputs to the forward models are not coherent with the auditory information. In particular, in the middle graph the head is in a idle position (random joint position and motor command equals to 0); in the right graphs, joint and motor values follow periodical head movements shifted in time, compared to the actual auditory signal. In each of the three tests, the upper plots show the Mel log-filterbank energies extracted from the original auditory signal. The plots in the central row show predicted log-filterbank energies, where the input state and motor information vary according to the test. The bottom plots show the output of the ego-noise suppression (predicted signal subtracted from the original one).

Kajić, I., Schillaci, G., Bodiroža, S., and Hafner, V. V. (2014). Learning hand-eye coordination for a humanoid robot using soms. In *Proc. of ACM/IEEE Int. Conf. on Human-robot Interaction (HRI2014)*, pages 192–193. ACM.

Knoblich, G. and Flach, R. (2001). Predicting the effects of actions: Interactions of perception and action. *Psychological Science*, 12(6):467–472.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69.

Maravita, A. and Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, 8(2):79 – 86.

Miikkulainen, R. (1990). *DISCERN: A Distributed Artificial Neural Network Model Of Script Processing And Memory*. PhD thesis, University of California.

Möller, R. and Schenck, W. (2008). Bootstrapping cognition from behavior — a computerized thought experiment. *Cognitive Science*, 32(3):504–542.

Morse, A. F., Greef, J. D., Belpaeme, T., and Cangelosi, A. (2010). Epigenetic robotics architecture (ERA). *IEEE Transactions on Automnomous Mental Development*, 2(4):325–339.

Pitti, A., Mori, H., Kouzuma, S., and Kuniyoshi, Y. (2009). Contingency perception and agency measure in visuo-motor spiking neural networks. *IEEE Transactions on Autonomous Mental Development*, 1(1):86–97.

Sahidullah, M. and Saha, G. (2012). Design, analysis and experimental evaluation of block based transformation in {MFCC} computation for speaker recognition. *Speech Communication*, 54(4):543 – 565.

Schillaci, G., Hafner, V., and Lara, B. (2012). Coupled inverse-forward models for action execution leading to tool-use in a humanoid robot. In *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2012)*, pages 231–232.

Schillaci, G., Hafner, V., and Lara, B. (2014). Online learning of visuo-motor coordination in a humanoid robot. a biologically inspired model. In *IEEE Int. Conf. on Development and Learning and Epigenetic Robotics (ICDL-Epirob2014)*, pages 130–136.

Schillaci, G., Hafner, V., Lara, B., and Grosjean, M. (2013). Is that me? sensorimotor learning and self-other distinction in robotics. In *ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2013)*, pages 223–224.

Weiss, C., Herwig, A., and Schütz-Bosbach, S. (2011). The self in action effects: Selective attenuation of self-generated sounds. *Cognition*, 121(2):207–218.

Wilson, M. and Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131:460–473.

Wolpert, D. M., Ghahramani, Z., and Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, 5(11):487 – 494.