ARTÍCULO

A Simple Habituation Mechanism for Perceptual User Interfaces *

O. Déniz, J. Lorenzo, M. Hernández
Universidad de Las Palmas de Gran Canaria
Instituto Universitario de Sistemas Inteligentes y
Aplicaciones Numéricas en Ingeniería
Edificio Central del Parque Científico-Tecnológico
Campus Universitario de Tafira
35017 Las Palmas - SPAIN
odeniz@dis.ulpgc.es

Abstract

Complex human-computer interfaces are more and more making use of high-level concepts extracted from sensory data for detecting aspects related to emotional states like fatigue, surprise, boredom, etc. Repetitive sensory patterns, for example, almost always will mean that the robot or agent will switch to a "bored" state, or that it will turn its attention to other entity. Novel structures in sensory data will normally cause surprise, increase of attention or even defensive reactions. The aim of this work is to introduce a simple mechanism for detecting such repetitive patterns in sensory data. Basically, sensory data can present two types of monotonous patterns: constant frequency (be it zero or greater than zero, be it a unique frequency or a wide spectrum) and repetitive frequency spectrum changes. Both types are considered by the proposed method in a conceptually and computationally simple framework. Experiments carried out using sensory data extracted both from the visual and auditory domains show the validity of the approach.

Keywords: Habituation, Novelty detection, Human computer interaction, Anthropomorphic robot.

1 Introduction

Living beings possess habituation mechanisms that allow them to ignore repetitive stimuli. If such stimuli were not gradually ignored, the continuous response would lead the living being to complete exhaustion. A large amount of information is continually received from the environment, and it has to be somehow filtered so that the agent can focus on the interesting data. Marsland [12] defines habituation as "a way of defocusing attention from features that are seen often". Many animals, and humans too, have some kind of mechanism to filter uninteresting stimuli.

*This work was partially funded by Canary Islands Regional Government *PI2000/042* and Universidad de Las Palmas de Gran Canaria *UNI2002/16* research projects.

Habituation is a filtering mechanism that has received a lot of attention in the physiological and psychological areas. In the physiological area, some researchers have investigated the mechanisms of habituation in animals, being one of the most known works the study of the Aplysia's gill-withdrawal reflex [2]. When the animal's siphon is touched, its gill contracts for a few seconds. If the siphon is stimulated repeatedly, the gill-withdrawal effect tends to disappear. Crook and Hayes [4] comment on a study carried out on two monkeys by Xiang and Brown who identified neurons that exhibit a habituation mechanism since their activity decreases as the stimulus is shown repeatedly.

Stanley's model [16] of habituation, proposed to simulate habituation data obtained from the cat spinal cord, has been widely used in the literature. This model describes the decrease efficacy y of a synapsis by the first-order differential equation:

$$\tau \frac{dy(t)}{dt} = \alpha(y_0 - y(t)) - S(t), \qquad (1)$$

where y_0 is the normal, initial value of y, $S(t)$ represents the external stimulation, τ is a time constant that governs the rate of habituation and α regulates the rate of recovery. Equation (1) ensures that the synaptic efficacy decreases when the input signal $S(t)$ increases and returns to its maximum y_0 in the absence of an input signal.

The model given by (1) can only explain short-term habituation, so Wang [19] introduced a model to incorporate both short-term and long-term habituation using an inverse S-shaped curve,

$$\tau \frac{dy(t)}{dt} = \alpha z(t)(y_0 - y(t)) - \beta y(t) S(t) \qquad (2)$$

$$\frac{dz(t)}{dt} = \gamma z(t)(z(t) - l) S(t), \qquad (3)$$

where α, y_0 and γ have the same meaning than in (1), β regulates the habituation and $z(t)$ decreases monotonically with each activation of the external stimulation $S(t)$, and models the long-term habituation. Due to this effect of $z(t)$ after a large number of activations, the recovery rate is slower.

Novelty detection is a concept related to habituation. Novelty detection is the discovery of stimuli not perceived before and so habituation serves as a novelty filter [17]. The rest of this paper is organized as follows. In Section 2 we briefly describe some applications of habituation mechanisms. In Section 3 the proposed method is described in detail. Experimental results are evaluated in Section 4. A brief discussion appears in Section 5. Finally in Section 6 we outline the main conclusions and future work.

2 Motivation

From a engineering viewpoint, perceptual user interfaces, like human-like robots, should be endowed with a habituation mechanism. The interest is twofold. First, it would be a filtering mechanism, discarding (or minimizing the importance of) repetitive information while paying attention to new experiences. This is in part motivated by the desire to distinguish between artificial and human signals. Artificial signals are often static or repeat with a fixed frequency. We do not want our robot to pay much attention to the hands of a wall-mounted clock. Instead, it would be more interesting to detect non-repetitive stimuli, such as a conversation or a sudden loud noise. Note that we generally consider monotonous signals as those having a fixed frequency or frequencies (which can be zero, that is, the signal does not change) but signals whose frequency changes in a periodic pattern could also be considered monotonous. Higher scales are also possible but we do not consider them in this work because they are very hard to visualize and real examples of them are not so common.

Second, habituation would lead to a more human-like behaviour, as perceived by users of the interface. As an example of this, consider the multimodal interface Kismet [1]. Someone can catch the eye of the system while waving a hand in its visual field of view, but if the stimulus is repetitive for a long time the system can show a lack of interest in it. Many aspects of Kismet's mental architecture are directly or indirectly influenced by the detection of monotonous sensory signals: stimulation and fatigue drives and the arousal dimension of its affect space (and in turn some emotional states, like surprise, boredom or interest).

Although we focus our work on the abilities described above, many other applications are also imaginable. In the robotics field, habituation mechanisms have been used to reduce oscillations caused by collision-avoidance behaviours when navigating through a narrow corridor [3]. Marsland [13] uses a SOM neural network as a memory

for novelty detection. To add short-term habituation to the original network, each neuron of the SOM is connected to an output neuron with habituable synapses based on the model (1). Habituation is also used in [18] for controlling reactivity strength, visual attention [15, 1], and general learning [5]. On the other hand, there is considerable interest in the field of musicology in Beat Tracking Systems (BTS) [8]. BTS systems aim to find the tempo of an audio signal, which is basically the rate of repetitions. The main applications of BTS systems are audio/video editing, synchronization of computer graphics with music, stage lighting control and audio content searching.

3 Proposed Method

If we use the model of Equation (1) we can obtain undesired effects with certain stimuli. A periodic input signal (with frequency greater than zero) can produce a response that does not exhibit habituation. This is due to the fact that the model does not account for changing stimuli, but for continuous ones. In order to include this fact in the model, we propose to use an auxiliary signal which will be zero when the stimulus is stationary or with a fixed frequency, and one otherwise, and use this signal as an input to the habituation model (1).

The auxiliary signal, which basically detects monotonous stimuli, is obtained from the spectrogram of the stimulus itself. The spectrogram is a time-frequency distribution of a signal, and it is based on the Fourier Transform with a sliding window [9]. The equation

$$\Phi(t,f) = \left| \int_{-\inf}^{inf} x(\tau) e^{(t-\tau)^2/T^2} e^{-j2\pi f\tau} d\tau \right|^2 \quad (4)$$

gives the definition of a spectrogram with a Gaussian window function of half-width T, and it is the power spectrum of a signal which corresponds to the squared magnitude of the Fourier transform of the windowed signal. The window can have other forms apart from the Gaussian one. In Figure 1 we show an audio signal and its corresponding spectrogram, where brighter areas correspond to higher power. Two well defined frequency spectra can be distinguished, for there is a change in the input signal at time 0.5 s. Temporal patterns of the stimulus signal have a specific pattern in the spectrogram. A fixed frequency signal corresponds to a straight line parallel to the time axis in the spectrogram, and the length of this line indicates how long has been the stimulus present.

Spectrograms are computed from windows of the input signal. These windows, of length l, overlap by $l-1$ samples. Let each spectrogram be represented as a matrix M, in which rows represent frequencies and columns represent time. We calculate the variance of each row of M, which produces a column vector $\mathbf{v}$. The norm of this vector $\mathbf{v}$ is a measure of how monotonous the input signal is. The norm will be high when the signal is changing, and low otherwise. Thus, the auxiliary signal needed is simply the thresholded norm of $\mathbf{v}$. The amplitude of the input signal affects the power content of the spectrograms, and in turn the norm of $\mathbf{v}$. Thus, prior to calculating the FFT the input signal must be normalized dividing each input window by the sum of its absolute values. A value of 1 for the auxiliary signal will mean that there are changes in the input signal, while a value of 0 indicates that the input signal is monotonous. Once the auxiliary signal is available, the model (1) is used to get the desired habituation behaviour, as controlled by parameters τ and α.

Formally, let N and l be the number of rows and columns of M, respectively, and let $m_{i,j}$ represent the element in row i and column j of M. Vector $\mathbf{v}$ is calculated as:

$$v_i = \frac{\sum_{j=1}^{l}(m_{i,j}-\mu_i)^2}{l} \;\; ; \;\; i = 1,\ldots,N \quad (5)$$

where:

$$\mu_i = \frac{\sum_{j=1}^{l} m_{i,j}}{l} \;\; ; \;\; i = 1,\ldots,N \quad (6)$$

The auxiliary signal is then, for a given threshold T:

$$A = \begin{cases} 1 & \text{if} \quad |\mathbf{v}| > T \\ 0 & \text{if} \quad |\mathbf{v}| \leq T \end{cases} \quad (7)$$

With this method both static an fixed frequency stimuli can be detected. However, there are stimuli that change their frequency according to a periodic pattern. These stimuli should also be considered as monotonous. The hissing sound of a

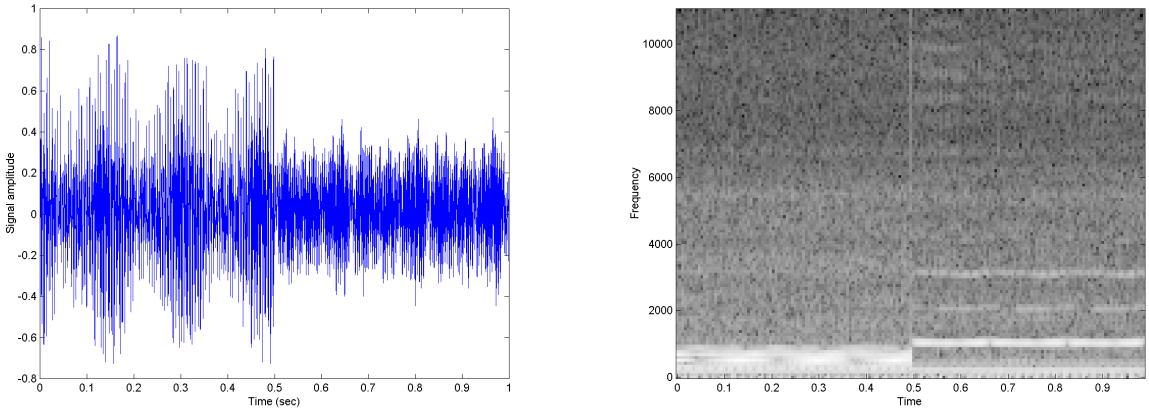


Figure 1: Audio signal (left) and its corresponding spectrogram (right).

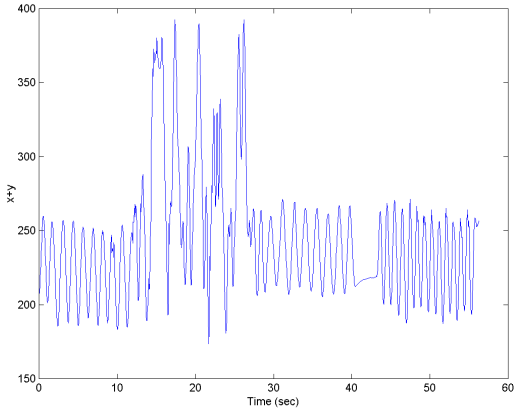
siren, for example, is a signal whose frequency changes in a repeated pattern. After few repetitions the signal will be considered monotonous. One way to detect these kind of stimuli is to use the same method with the auxiliary signal. If the input signal changes its frequency content in a repeated pattern, the auxiliary signal will be periodic with a fixed frequency, and that can be detected as explained in the previous paragraph. Thus, two thresholds will be needed, one for the "first level" and one for the "second level". Higher levels could conceivably be used, but we have not considered them because they are very difficult to visualize and encounter in the physical world. Note that the second-level auxiliary signal will be 1 when there are changes in the first-level auxiliary signal, and thus when there are changes in the input signal, and 0 otherwise. Thus, the final input to the habituation model (1) will be the second-level auxiliary signal. Note that this second level introduces additional computation, and in some cases we could consider it unnecessary, if we decide to detect only simple monotonous signals.

There is only one detail left. If the first-level auxiliary signal is 1 (meaning that the input signal is changing), and this remains for a while, the second-level auxiliary signal will be 0 (because the second-level norm of the variance vector will be 0) which is not the correct value. In order to correct this, the second level must detect when the norm is 0 and, if so, use the value of the first-level auxiliary signal, instead of the second-level auxiliary signal. Note that if the first-level auxiliary signal is periodic the second-level variances obtained should theoretically be 0, which would prevent the use of this correction. However, in all the experiments carried out this never happened, because there is always an unavoidable amount of fluctuations in the input signal, which makes the variances larger than 0.

A previous version of the method proposed here has been already published elsewhere [11, 10]. That version used only the frequency associated to the maximum power. Habituation should be present when the plot of that frequency versus time is a straight line. Changes are detected by fitting a line to the last k values of the frequency and computing the difference between the current value and the predicted value with the fitted line. That approach, however is too simplistic in the sense that it assumes that the input signal is entirely represented by the frequency of maximum power.

4 Experiments

The algorithm described in Section 3 was implemented to test it with different input signals. The first experiments that we present use only the first level mentioned in Section 3. In order to gather signals from the visual domain, we recorded video containing a yellow bright stimulus (a yellow card) that was moved in a repetitive fashion, see Figure 2-a). Using simple segmentation techniques we extracted the centroid of the card on each frame (384x288) and summed the x and y pixel coordinates to form the one-dimensional signal of Figure 2-b). The sequence of card movements throughout the recording was: horizontal movement, random (aperiodic) movement, vertical movement and vertical movement



a) b)

Figure 2: a) Video recording used for the visual habituation experiment, b) one-dimensional signal extracted from it.

at a different frequency than the previous one.

The results appear in Figure 3. Windows of 128 samples were used, and the variance threshold was set at 1000.

As for the audio domain, we recorded signals with a standard PC microphone, at a 22050 Hz sample rate, 8 bits. Figure 4 shows the results obtained for an audio signal that contains three sequential parts: silence (0-0.5s), people speaking (0.5-1s) and a tone played from an electric piano (1-1.4s). Note that there is an initial delay due to the need to fill the input window, here of length $l = 5120$. The habituation level, obtained using the model of (1), shows a satisfactory response.

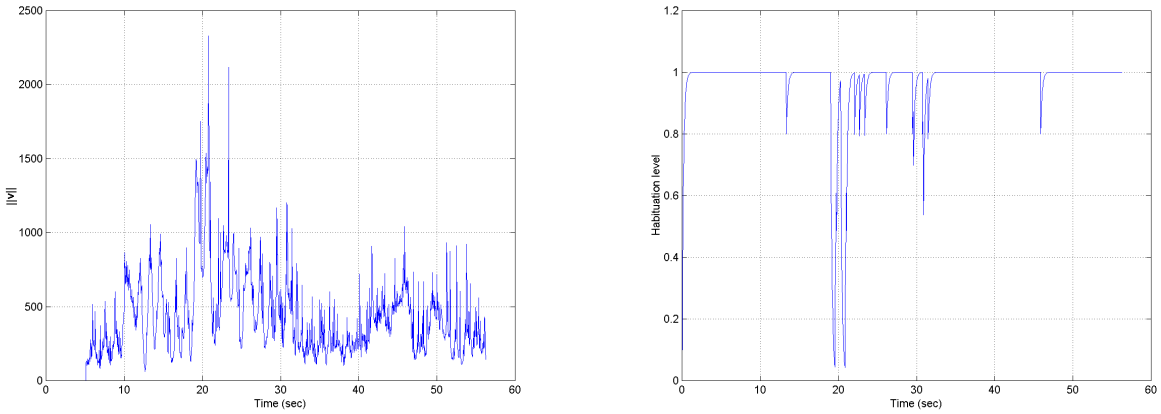
Figure 5 shows the results obtained for an audio signal that contains another three sequential parts: a tone played from an electric piano (0-0.5s), silence (0.5-1s) and another tone (1-1.4s). The same window length $l = 5120$ was used, and again the habituation level shows a satisfactory behaviour.

In order to test both the first and second levels of the method, we built an audio signal containing three sequential parts: a beep repetitive sound from a mobile phone, people speaking and a tone played from an electric piano. This signal was accelerated to reduce computation time, which does not alter the qualitative results of the experiments. Results are shown in Figure 6. The window length was $l = 5120$ for the first level and $l = 2148$ for the second. In this case the repetitive beeps (clearly observed as a repetitive pattern in the first part of the spectrogram) are correctly considered as monotonous. This would not have occurred if we had used the first-level auxiliary signal alone, for numerous changes are detected (see Figure 6-d).

5 Discussion and Implementation

In this section we discuss a few aspects of practical interest. Particularly, we will comment on the effect of the values of the different parameters to use:

- Length of the input window, l: It should be the largest possible, in order to detect stimuli with large period. However it cannot be too large because that would introduce an unacceptable delay in the response to stimuli with smaller period. Thus, it depends on the type of stimuli. A flexible solution would be to implement multiple instances of the problem, each one with a different size for this parameter, in a multiscale fashion.
- Tau, τ: It controls the rate of habituation.
- Alpha, α: It controls the rate or recovery.
- Number of discrete frequency levels, N: Dependent on the type of input stimulus, it should normally be the largest possible. For the case of auditive signals, the minimum noticeable difference that people can distinguish is as low as 1.3Hz [14]. Other input



a) b)

Figure 3: a) Evolution of the (l_2) norm of the variance vector $\mathbf{v}$, b) habituation level, using $\tau = 5, \alpha = 1$.

stimuli could be sonar data, blob positions, pressure readings, etc.

- Variance thresholds: They refer to the minimum change in the frequency spectrum to detect a change in the signal. If set too high, we run the risk of ignoring signal changes. If set too low, "hypersensitive" responses could be obtained. The appropriate values depend both on the type of input signal and the number of discrete frequency levels. These thresholds could be changed depending on the amount of available resources. If available resources are high, a lower threshold could be appropriate (producing more sensitivity or attention). Otherwise, a higher threshold would produce a more believable response.

The computational cost of the proposed method is basically dependent on the calculus of the spectrogram. This, in turn, basically depends on the FFT. Thus, the total cost, for a window of length l is $l \log_2 l$ (for the first-level alone). This is therefore the cost of producing a new value of the auxiliary signal for each input sample. If a multiscale (multiple values for l) approach is used, the multiple instances of the problem can use parallel computation. Also, the second-level part of the problem can be solved in parallel with the first-level.

The habituation mechanism described here has been implemented for an anthropomorphic robotic head that is being developed at our laboratory [7]. The robotic head is intended as a multimodal interface with human-like abilities. The habituation mechanism has been implemented for signals in the visual domain, i.e. images taken by the cameras placed in the eyes. The difference between the current and previous frame is calculated. Then it is thresholded and filtered with the Open and Close operators. Also, blobs smaller than a threshold are removed. Then the center of mass of the resultant image is calculated. The signal that feeds the habituation algorithm is the sum of the x and y components of the center of mass. When the image does not show significant changes or repetitive movements are present for a while the habituation signal grows. When it grows larger than a threshold, an inhibition signal is sent to the attention module of the robot [6], which then changes its focus of attention. The head pan and tilt movements produce changes in the images, though it was observed that they are not periodic, and so habituation does not grow.

6 Conclusions

Habituation is the decrease in the strength of an agent's response when it receives repetitive stimuli. This ability is present in almost any living being, and it is of capital importance, if we consider the effect that a lack of it would have. In this work we propose a simple spectrogram-based algorithm for detecting monotonous input signals, independent of their sensory origin (auditive, visual, ...). Signals that repeat with constant frequency or frequencies are considered monotonous. Signals that present a periodic changing pattern in their frequency content can also be considered monotonous. The usefulness of the algorithm is

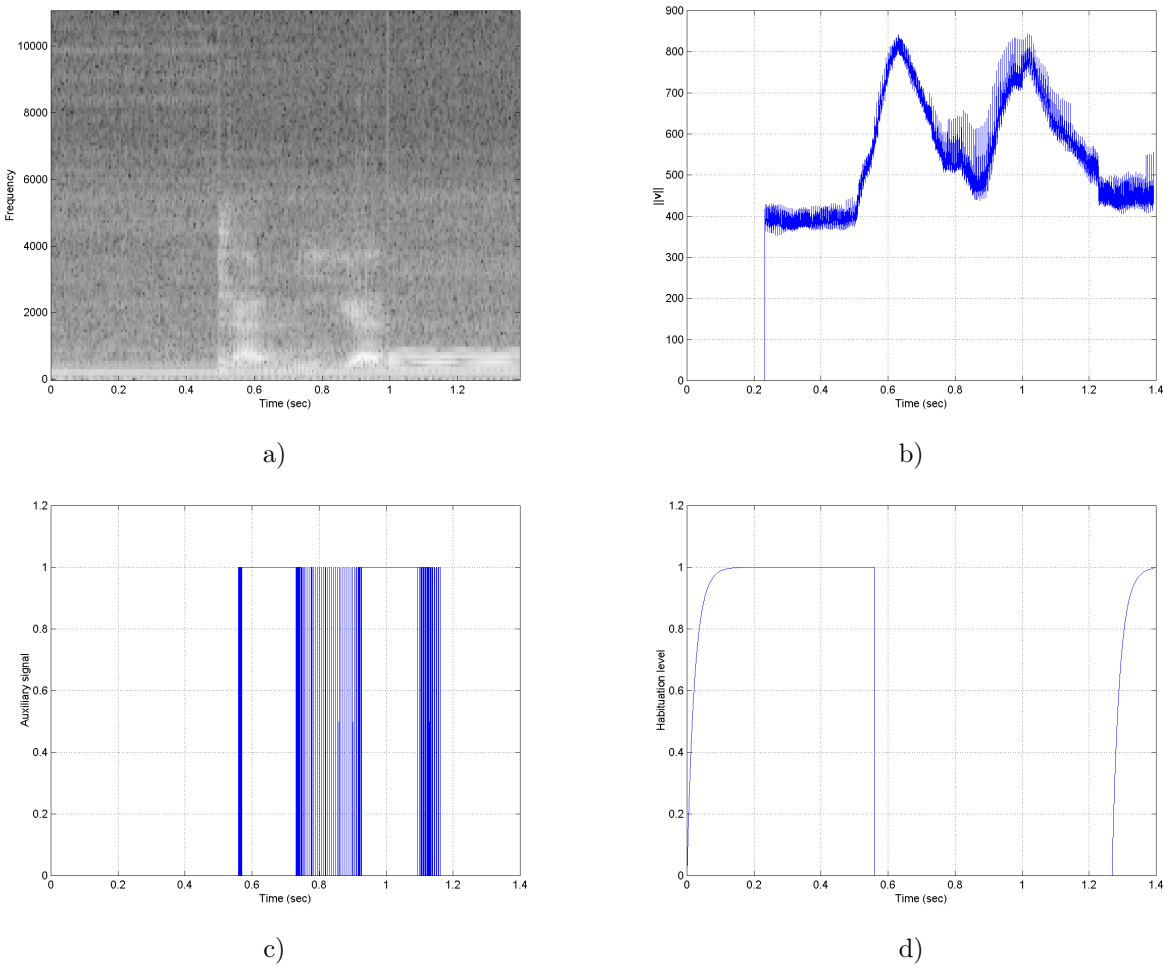


Figure 4: a) Spectrogram of the audio signal, b) evolution of the (l_2) norm of the variance vector $\mathbf{v}$, c) auxiliary signal, obtained using a threshold of 600, d) habituation level, using $\tau = 1, \alpha = 0.002$.

evaluated in experiments with signals gathered both from the visual and the auditive domains.

References

[1] C. Breazeal and B. Scassellati. How to build robots that make friends and influence people. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1999. Korea.

[2] V. Castellucci, H. Pinsker, I. Kupfermann, and E. R. Kandel. Neuronal mechanisms of habituation and dishabituation of the gill-withdrawal reflex in *Aplysia*. *Science*, 167:1745–1748, 1970.

[3] C. Chang. Improving hallway navigation in mobile robots with sensory habituation. In *Proc. of the Int. Joint Conference on Neural Networks (IJCNN2000)*, volume V, pages 143–147, Como, Italy, 2000.

[4] P. Crook and G. Hayes. A robot implementation of a biologically inspired method ofr novelty detection. In *Proceedings of TIMR 2001 - Towards Intelligent Mobile Robots*, Manchester, April 2001.

[5] R. Damper, R. French, and T. Scutt. Arbib: an autonomous robot based on inspirations from biology, in press, 1999.

[6] O. Déniz, M. Castrillón, and M. Hernández. Multimodal attention system for an interactive robot. In *Lectures Notes in Computer Science. 1st Iberian Conference on Pattern Recognition and Image*, volume 2652, June 2003. Mallorca, Spain.

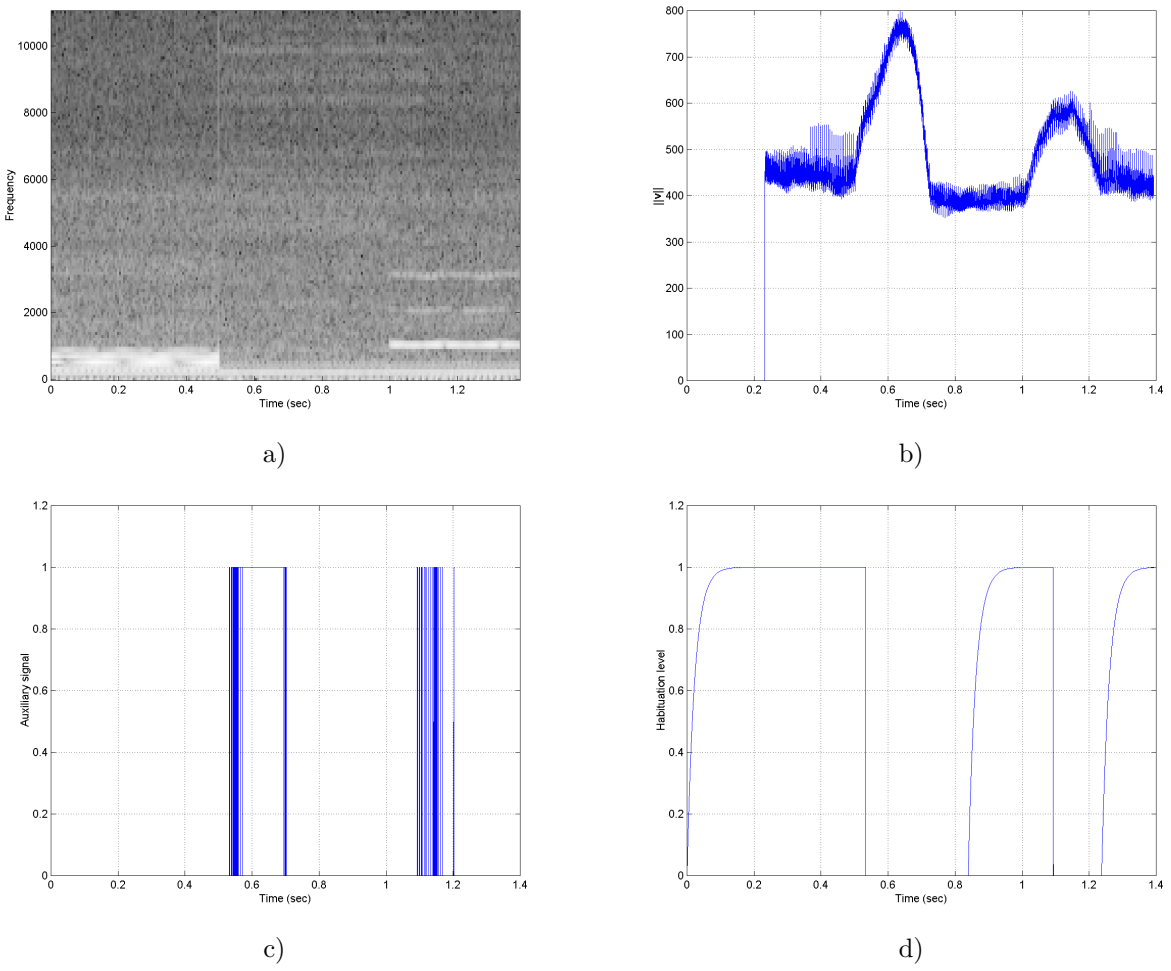


Figure 5: a) Spectrogram of the audio signal, b) evolution of the (l_2) norm of the variance vector $\mathbf{v}$, c) auxiliary signal, obtained using a threshold of 600, d) habituation level, using $\tau = 1, \alpha = 0.002$.

[7] O. Déniz, M. Castrillón, J. Lorenzo, C. Guerra, D. Hernández, and M. Hernández. CASIMIRO: A robot head for human-computer interaction. In *Proceedings of 11th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN'2002)*, 2002. Berlin, Germany.

[8] M. Goto and Y. Muraoka. Issues in evaluating beat tracking systems. In *Working Notes of the IJCAI-97 Workshop on Issues in AI and Music, (in press)*, 1997.

[9] S. Holland, T. Kosel, R. Waver, and W. Sachse. Determination of plate source, detector separation fron one signal. *Ultrasonics*, 38:620–623, 2000.

[10] J. Lorenzo and M. Hernández. Habituation based on spectrogram analysis. In *Lecture Notes in Artificial Intelligence. Advances in Artificial Intelligence, IBERAMIA 2002*, volume 2527, 2002. Sevilla, Spain.

[11] J. Lorenzo and M. Hernández. A habituation mechanism for a perceptual user interface. In *Proceedings of 11th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN'2002)*, 2002. Berlin, Germany.

[12] S. Marsland, U. Nehmzow, and J. Shapiro. A model of habituation applied to mobile robots. In *Proceedings of TIMR 1999 - Towards Intelligent Mobile Robots*, Bristol, 1999.

[13] S. Marsland, U. Nehmzow, and J. Shapiro. Detecting novel features of an environment using habituation. In *From Animals to Animats, The Sixth International Conference*

on Simulation of Adaptive Behaviour, Paris, 2000.

[14] Time and Frequency Division, Physics Laboratory, NIST. Time and frequency from A to Z. Available at `http://www.boulder.nist.gov/timefreq/general/glossary.htm`.

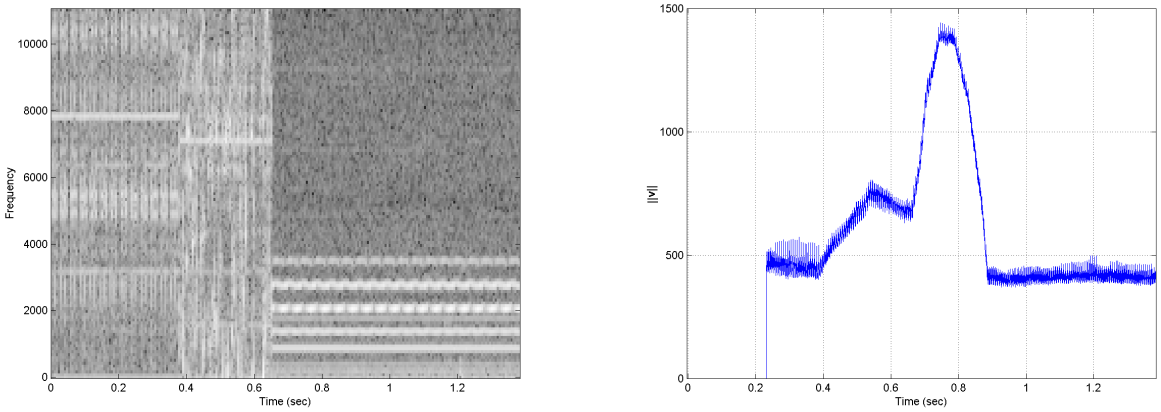
[15] M.W. Peters and A. Sowmya. Integrated techniques for self-organisation, sampling, habituation, and motion-tracking in visual robotics applications. In *Proceedings of Machine Vision Applications '98*, 1998. Tokyo, Japan.

[16] J.C. Stanley. Computer simulation of a model of habituation. *Nature*, 261:146–148, 1976.

[17] B. Stiles and J. Ghosh. A habituation based neural network for spatio-temporal classification. In *Proceedings of the 1995 IEEE Workshop In Neural Networks for Signal Processing*, pages 135–144, Cambridge, MA, September 1995.

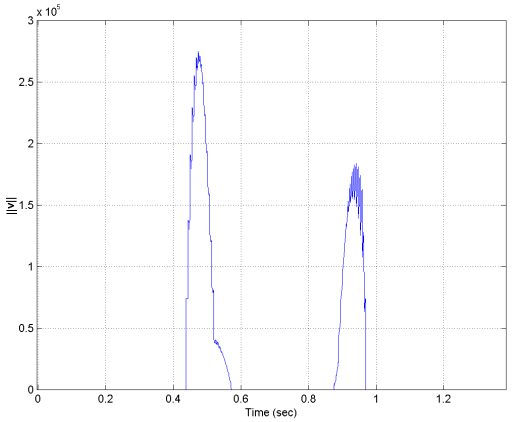
[18] A. Stoytchev and R.C. Arkin. Combining deliberation, reactivity and motivation in the context of a behavior-based robot architecture. Available at `http://www.cc.gatech.edu/ai/robot-lab/publications.html`.

[19] D. Wang. Habituation. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 441–444. MIT Press, 1995.

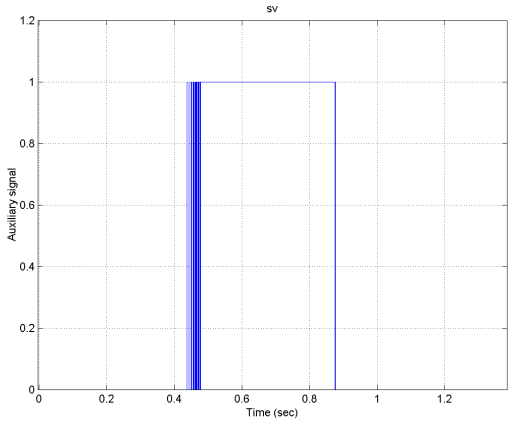


a)

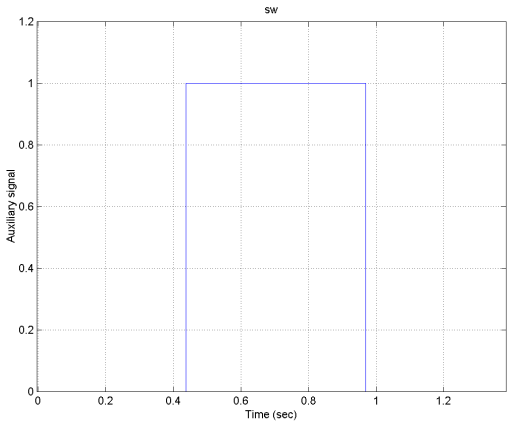
b)



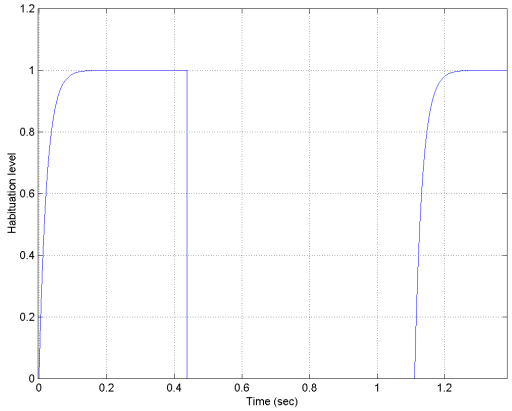
c)



d)



e)



f)

Figure 6: a) Spectrogram of the audio signal, b) evolution of the first-level (l_2) norm of the variance vector, c) evolution of the second-level (l_2) norm of the variance vector, d) first-level auxiliary signal, obtained using a threshold of 600, e) second-level auxiliary signal, obtained using a threshold of 1000, f) habituation level, using $\tau = 1, \alpha = 0.002$.