Investigating and modeling the effects of cognitive factors on multisensory perception using skin conductance

Patricia Besson*, Lionel Bringoux*, Stéphanie Giraud* and Christophe Bourdin*
*CNRS & Aix Marseille Université
UMR 7287 Institute of Movement Sciences, 13288 Marseille, France
Email: patricia.besson@univ-amu.fr

***Abstract*—Audiovisual perception associated to a spatial localization task is investigated using a focused attention paradigm: in presence of both acoustic and visual stimuli, subjects are required to localize either one or the other of these stimuli. Behavioral measures (i.e. subjects' localization errors), as well as skin conductance responses (i.e., components of the orienting responses) are acquired during the experiment. The subjects' performance on the localization task differ depending on the sensory nature of the stimulus they have to localize. The processing of the incoming information and the mobilized cognitive resources also depends on the task, as suggested by the analysis of the skin conductance responses. The latter are incorporated to a bayesian network inferring the subjects' error given audiovisual stimulus positions. This amounts to provide the model with some knowledge about the cognitive factors that interface with multisensory perception (resulting in a different exploitation of similar audiovisual information). As a result, the subjects' performance are better estimated by the model.**

I. Introduction

Human beings are continuously interacting with their environment. This interaction can be characterized in term of successive perception-action loops: the information received through the senses are interpreted and organized in order to build a coherent representation of the environment, used thereafter to make decisions and to undertake actions adapted to the individuals intentions.

Therefore, actions are not based on the reality itself, but on an interpretation of this reality, brought by perception. Perception is multisensory, which means that each sensory modality yields information, used in interaction to build a unied perceptual experience. Understanding these interactions could provide important key in designing effective interfaces, or assistance systems [1]–[3]. Hence, the problem of multisensory integration is an active field of research. Multisensory perception can be understood as a maximum likelihood estimation process, whose purpose is to come up with a multisensory estimate more reliable than the individual estimates (see e.g. [4], [5]). In this perspective, it amounts essentially to a bottom-up process. However, recent works have shown that cognitive (top-down) processes also step in multisensory perception and should be taken into account by the model (see [6] for a review).

In [7], Kohler et al. investigated the role of intention and attention in the perception of an ambiguous movement, stressing the effect of voluntary control and attentional focus on perception. Also, in previous works, we have established, using information theory and bayesian networks (BN), that subjects use a same audiovisual information differently depending on the sensory nature (i.e., acoustic or visual) of the stimulus they are instructed to localize [8], [9]. Therefore, the goal and expectations of the subjects are some of the cognitive factors modulating multisensory perception. This modulation is certainly related to some extent to attentional mechanisms, that prevent perceptual overload through selection mechanisms (that are also bottom-up or top-down) interfacing multisensory perception [10]. It is still an open question to understand how multisensory integration and crossmodal attention are related [2], [10].

Attentional mechanisms have been largely investigated through the study of the orienting response, assumed as an involuntary attentional mechanism that alerts the organism when novel and significant stimuli occur [11]–[13]. Information processing theories of the orienting response suggest that it is associated to the amount of resources allocated to the processing of a stimulus [14]. A widely used component of the orienting response is the skin conductance response (SCR) [15], [16]. Indeed, electrodermal activity is mainly under the control of the sympathetic nervous system [15], responsible for mobilizing the organism resources.

The purpose of this study was to further investigate the interplay between cognitive factors and audiovisual perception in a spatial localization task, using a paradigm similar to the one presented in [8], [9]. Shortly, subjects are exposed to acoustic and visual stimuli that are temporally but not necessary spatiallay coincident, and must localize either the acoustic or the visual stimulus, depending on the instruction they received. As already mentioned, we mathematically established in [8], [9] the difference in processing the audiovisual information depending on the given instructions. Our hypothesis was that these different processings of the incoming information, likely to be related to different attentional focus, should yield different orienting responses. Therefore, we add to the measure of behavioral features (i.e., subjects' pointing errors), the analysis of the SCR, as a component of the orienting response. Bayesian network models are proposed to investigate the results of the experiment from a computational point of

view, the objective being to define a model of multisensory perception where cognitive factors are acknowledged.

Sec. II details the experimental protocol. The analysis of the experimental data is carried out in sec. III, and performance of Bayesian network models inferring the subjects' localization errors using or not the SCR as input are presented. The results of both experiment and model are discussed in sec. IV.

II. Method

A. Subjects

Four right-handed participants (mean age: 25.8 years) took part in the experiment. They were free from any auditory or visual defect (as attested by classical audiograms or sight tests carried out before the experiment). All participants gave informed consent prior to the study, according to Aix-Marseille University regulations and the 1964 Declaration of Helsinki. They were nevertheless naïve as to the purpose and the manipulated factors of the experiment.

B. Material

Participants were seated in complete darkness in an anechoic audiovisual stimulation room. The 80ms long visual stimuli were generated by five red light-emitting diodes (LEDs) arranged horizontally at eye level, along the arc of a circle of raidus 57.5 cm attached to the chair, in front of the subject. The central LED was adjusted on the cyclopean eye of each seated subject, the 4 other LEDs being positioned at $\pm 10^\circ$ and $\pm 20^\circ$ from this central target. In addition, a white noise emitter tweeter located just above this diode trail could be moved circularly by the experimenter and be placed at the same positions than the visual targets. This tweeter emitted 80ms of white noise that defined the acoustic stimuli. No prior inspection of the setup was made available to the subjects.

C. Procedure

The experiment was divided into two counterbalanced sessions, during which participants were systematically exposed to temporally synchronous visual and auditory stimuli, and had to judge the position of the primary stimulus, that could be either the auditory stimulus (i.e., acoustic session) or the visual stimulus (i.e., visual session). To that aim, participants were required to orient a fixed-base pointer connected to a potentiometer towards the auditory or visual targets, depending on the session. In both sessions, synchronous visual and auditory stimuli were either spatially congruent or not. In the latter case, the spatial mismatch between the stimulus positions was 20°. A session was made of 300 trials, where congruent and non-congruent stimuli were pseudo-randomly presented with a 50% probability.

In both tasks, the displacement and the final location of the pointer, corresponding to the perceived auditory or visual target position, were recorded. The skin conductance (measured with electrode pairs positioned on the distal phalanges of the left index and little fingers) was also recorded during both sessions and was temporally related to each judgement for subsequent analyses.

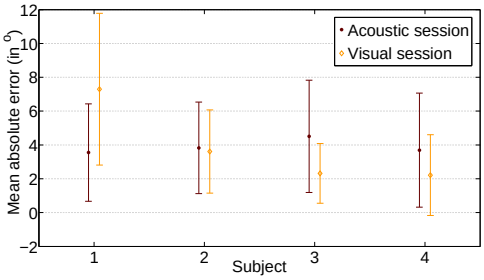

Fig. 1. Mean absolute subjects' errors (Δ) when localizing the acoustic or visual stimuli. The error bars stand for the associated standard deviations.

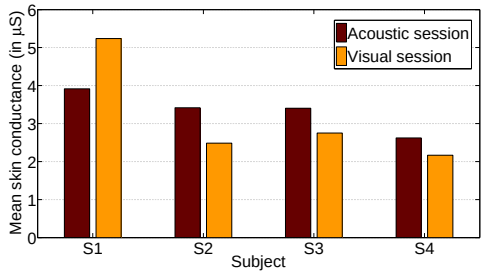

Fig. 2. Average subjects' mean skin conductance responses (SCRs) on each trial, during the acoustic and visual sessions.

Notice that two unisensory sessions (where only acoustic or visual stimuli occurred) of 25 trials were also carried out in order to get the subjects' intrinsic errors in acoustic or visual stimulus localizations.

III. Analysis

A. Statistical analysis of the results

First of all, for each session, the subjects' pointed positions were normalized by substracting their mean errors on the unimodal tasks. Though the experimental conditions (use of an anechoic room) permitted to produce acoustic stimuli of high quality, we expected the subjects to be better in localizing the visual stimuli (visual capture). Fig. 1 shows the mean and standard deviation values of the subjects' absolute errors Δ (absolute value of the difference between the pointed position and the primary stimulus position), when localizing the acoustic and visual stimuli. Three subjects (subjects 2, 3 and 4) out of four were indeed more accurate and precise when localizing the visual stimuli rather than the acoustic ones. Surprisingly, it was the opposite for subject 1 (accuracy: 3.6° in the acoustic session, 7.3° in the visual session; precision: 2.9° and 4.5° in the acoustic and visual sessions respectively).

The emission of the stimuli yielded a skin conductance response, that decreased during the session (habituation effect). The mean value of the SCR on each trial was analyzed. The average value of these subjects' mean SCRs is shown on 2, for the acoustic and visuals sessions. It is larger in sessions where the average Δ is larger too, i.e., in the acoustic sessions for subjects 2, 3 and 4, and in the visual session for subject 1. It is also worth noticing the high correlation between differences

of pointing errors in the acoustic and visual session and differences of SCRs in these two sessions (pearson correlation coefficient $\rho = .89$).

Statistical tests were performed with the Matlab 7.6 Software and globally confirm the observations based on the visual inspection of the data. The null hypothesis stating that subjects' absolute errors on the acoustic and visual sessions come from the same statistical population was tested using Kruskall-Wallis tests, and can be rejected for each subject ($p < 0.001$) but for subject 2 ($p = 0.497$). Kruskall-Wallis statistical analyses of SCR mean values per trial also pointed out a significant difference between the acoustic and visual sessions for each subject ($p < 0.001$).

B. Model definition

These results state that subjects use a same available audiovisual information differently, depending on the task they are instructed to perform (localization of either the acoustic or visual stimuli), as pointed out in [8], [9]. By adding the analysis of subjects' SCRs to the analysis of subjects' performance carried out in [8], [9], we observe that the two tasks require the subjects to mobilize the organism in different ways (i.e., the response of the autonomic nervous system, therefore, the SCRs, differs). For each of the subject, one of the two tasks seems to be more difficult and requires a higher mobilization of the organism. Despite this higher mobilization, their performance on the task is lower (though the difference is not significant for Subject 2).

It would be interesting to test whether prediction of the subjects' performance can benefit from some knowledge about the orienting response elicited by the task. Indeed, in [9], we proposed a BN model that inferred the subjects' judgement from the positions of the emitted stimuli (bottom-up information). We showed that introducing in the model a rv N that stood for the sensory nature of the stimulus to be localized (i.e., the instructions received by the subjects and modulating their objectives, hence, cognitive factors interfacing multisensory perception) changed the structure of the BN model. Now, we want to investigate whether providing the model with undirect knowledge about these cognitive factors, in the form of the skin conductance component of the subjects' orienting responses, would give means to the model for discrimating between different ways of handling the same available information.

To this end, we tested in turn two BN models. In a first one, $\mathcal{M}_1$, random variables (rvs) modeling the positions S_1 and S_2 of the primary and secondary stimuli were the inputs of the model, which infers the subjects' absolute localization error Δ. In the second model, $\mathcal{M}_2$, the subjects' mean skin conductance response SCR was also used as input for the model to predict Δ[1]. The two models are shown on Fig. 3.

The probability density functions (pdfs) of the rvs S_1 and S_2 were estimated using multinomial approaches, both of them taking on values on the set $\{0, \pm 10, \pm 20\}$. The subjects' absolute errors were first normalized between 0 and 1, by taking the minimal and maximal errors on the experiment. They were modeled by the rv Δ, whose pdf is estimated using histogram of bins 0.2 width. The subjects' mean skin conductance responses SCR were also normalized between 0 and 1 for each subject, using the minimal and maximal values found on the whole experiment. The pdf of the rv SCR was estimated using histograms of bins 0.1 width.

[1]For simplification purpose, the rvs are named using the same acronyms than the signal they stand for.

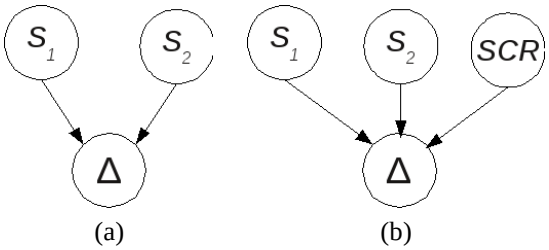

Fig. 3. Bayesian networks used to infer the subjects' absolute error when localizing the primary stimulus S_1. (a) without the mean SCR (model $\mathcal{M}_1$); (b) with the mean SCR as input (model $\mathcal{M}_2$).

TABLE I
MODELS' ERRORS E_1 (MODEL $\mathcal{M}_1$, WITHOUT SCR) AND E_2 (MODEL $\mathcal{M}_2$, WITH SCR). THE NORMALIZED DIFFERENCE BETWEEN THE MODELS' ERRORS ARE ALSO GIVEN.

Subject	1	2	3	4
E_1	0.1611	0.2374	0.1331	0.1080
E_2	0.1533	0.2124	0.1317	0.1068
$(E_2 - E_1)/E_1$	-4.84%	-10.53%	-1.05%	-1.11%

C. Model performance

The models were trained using a leave-one-out cross-validation scheme [17]: training was performed using all the observations but one, and testing on the remaining value, in turn over the 600 observations[2] of the dataset. Both the learning and inference stages were implemented using the Bayes Net Toolbox for Matlab [18]. The models' errors are defined as $E = |\Delta^* - \Delta|$, where Δ^* is the subject's absolute error inferred by the model and Δ the observed absolute error. E_1 is the mean error of model $\mathcal{M}_1$ and E_2 the mean error of model $\mathcal{M}_2$.

The models' performance are summarized in Table. I. Adding SCR to the model increases its performance: E_2 is 10% lower than E_1 for subject 2, and almost 5% lower for subject 2. For subjects 3 and 4, the gain is not as high as could be expected (maybe because of the use of very simple SCR features), though performance of model $\mathcal{M}_2$ still improve as compared to model $\mathcal{M}_1$.

IV. DISCUSSION

Multisensory perception refers to the exploitation and interpretation of the sensory information received through our multiple sensory captors. Different sensory percepts can be produced in presence of the same exogeneous stimulation, due to bottom-up factors (related to stimulus properties), but also,

[2]Two sessions of 300 trials.

top-down (cognitive) factors (knowledge, expectation, goals of the individual), that step in the process and should be taken into account by models of multisensory perception. This paper addresses the problem of accounting for cognitive factors in multisensory perception.

In the present experiment, subjects were required to make different use of the similar audiovisual information they received: in two different sessions, they had to localize either the acoustic or the visual stimulus, while both stimulus modalities were synchronously presented, in either conguent or non-congruent positions. The subjects' performance are not the same over the two sessions. According to maximum likelihood estimation model of multisensory perception [4], [5], the integration of multisensory information aims at optimizing the reliability of the integrated percept. Therefore, bottom-up factors will tend to favor the use of the more reliable source of information for the task at hand, that is, usually, visual information for a spatial localization task. Actually, three subjects out of four unsurprisingly reached better performance in the visual than in the acoustic localization task, but one subject (subject 1) shows the opposite performance scheme. Analyzing the subjects' field dependence or independence might have provided us with some explanations about this point (subject 1 seems to spatially localize stimuli with a higher reliability using the acoustic information).

Because of the randomness of the stimulus spatial congruency, subjects could not try to take advantage of the audiovisual information to perform the task, but should indeed focus on the stimulus of the primary modality (modality to be localized). This scheme can be understood as a focused attention paradigm [19]. One of the two tasks certainly impedes more than the other the subjects' "natural" way of processing the information: they will have to specifically drive their attention away from the modality they preferably rely on for a spatial localization task, in order to conform to the received instructions.

The process of organizing and interpreting the sensory inputs requires cognitive resources [20], that we can hypothesize to be different between the two tasks. To investigate this hypothesis, we recorded the skin conductance component of the orienting response, presumed to reflect the processing of incoming information through the related changes in the autonomic nervous system [21], [22]. The observations made on the SCR acquired during the acoustic and visual sessions confirm a different level of organism mobilization depending on the tasks: the mean SCRs on trials where subjects try to localize the acoustic stimuli differ from the mean SCRs on trials where they aim at localizing the visual one. Moreover, the higher level of cognitive resources mobilized by the subject (as reflected by higher mean SCRs) to perform the – presumably – more difficult task do not prevent the performance to decrease (positive pearson correlation coefficient between the performance and the mean SCRs).

We tried to exploit this interesting relationship between resource level and performance in a simple naïve Bayesian model. Indeed, a model trying to infer the subjects' errors from knowledge of the stimulus positions solely cannot predict the subjects' performance differences due to various subjects' objectives or focus of attention, i.e., due to the interference of cognitive factors. In [9], these cognitive factors were introduced in the model through a rv modeling the instructions received by the subject. As a result, the structure of the BN model changed depending on the sensory nature of the stimulus to be localized. In this paper, we propose a BN that takes as inputs not only the audiovisual stimulus positions, but also the mean SCR on each trial: this model achieves better performance than a BN where this SCR information is not available. These results show that adding to the model some knowledge about the subjects' orienting response, through the SCR, helps the model to discriminate between different performance schemes. That is, it provides the model with some clues about these different ways of processing a similar incoming sensory information, depending on hidden factors related to different subjects' objectives (cognitive factors).

V. Conclusion

We do think that these results could be used with advantage in human factor engineering, for the design of efficient interactive systems. However, these are only preliminary results, that require to be tested on larger subject sets prior to be robustly established. Also, the improvment of the model's performance obtained when taking SCR into account is not as high as what could be expected. Optimizing the features extracted from the skin conductance signal (amplitude and slope of the SCRs instead of the mean value for example) as well as including temporal information about the SCR signal (habituation characteristic in particular) and subjects' error dynamics, should lead to better results.

Acknowledgment

The authors would like to thank Alain Donneaud for his help in building the experimental device.

References

[1] C. D. Wickens and C. M. Carswell, "Information processing," in *Handbook of Human Factors and Ergonomics*, G. Salvendy, Ed. Hoboken, NJ, USA: John Wiley & Sons, Inc., Feb. 2006, pp. 111–149.

[2] T. Koelewijn, A. Bronkhorst, and J. Theeuwes, "Competition between auditory and visual spatial cues during visual task performance," *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, vol. 195, no. 4, pp. 593–602, June 2009, PMID: 19436999.

[3] C. Spence, "Crossmodal spatial attention," *Annals of the New York Academy of Sciences*, vol. 1191, no. 1, pp. 182–200, Mar. 2010.

[4] M. S. Landy, L. T. Maloney, E. B. Johnston, and M. Young, "Measurement and modeling of depth cue combination: in defense of weak fusion," *Vision Research*, vol. 35, no. 3, pp. 389–412, Feb. 1995, PMID: 7892735.

[5] M. O. Ernst and H. H. Bülthoff, "Merging the senses into a robust percept," *TRENDS in Cognitive Sciences*, vol. 8, no. 4, pp. 162–169, 2004.

[6] D. Alais, F. N. Newell, and P. Mamassian, "Multisensory processing in review: from physiology to behaviour," *Seeing and Perceiving*, vol. 23, no. 1, pp. 3–38, Mar. 2010.

[7] A. Kohler, L. Haddad, W. Singer, and L. Muckli, "Deciding what to see: the role of intention and attention in the perception of apparent motion," *Vision Research*, vol. 48, no. 8, pp. 1096–1106, Mar. 2008.

[8] P. Besson, J. Richiardi, C. Bourdin, L. Bringoux, D. R. Mestre, and J. Vercher, "Bayesian networks and information theory for audio-visual perception modeling," *Biological Cybernetics*, vol. 103, no. 3, pp. 213–226, May 2010.

[9] P. Besson, C. Bourdin, and L. Bringoux, "A comprehensive model of audiovisual perception: Both percept and temporal dynamics," *PLoS ONE*, vol. 6, no. 8, p. e23811, Aug. 2011.

[10] D. Talsma, D. Senkowski, S. Soto-Faraco, and M. G. Woldorff, "The multifaceted interplay between attention and multisensory integration." *Trends Cogn Sci*, vol. 14, no. 9, pp. 400–410, Sep 2010.

[11] Y. N. Sokolov, "Orienting reflex as information regulator," in *Psychological research in USSR*, A. Leontiev, A. Luria, and Smirnov, Eds. Moscow: Progress, 1966, vol. 1, pp. 334–360.

[12] J. A. Spinks, G. H. Blowers, and D. T. Shek, "The role of the orienting response in the anticipation of information: a skin conductance response study," *Psychophysiology*, vol. 22, no. 4, pp. 385–394, July 1985, PMID: 4023149.

[13] I. Gati, G. Ben-Shakhar, and S. Avni-Liberty, "Stimulus novelty and significance in electrodermal orienting responses: the effects of adding versus deleting stimulus components," *Psychophysiology*, vol. 33, no. 6, pp. 637–643, Nov. 1996.

[14] D. L. Filion, M. E. Dawson, A. M. Schell, and E. A. Hazlett, "The relationship between skin conductance orienting the allocation of processing resources," *Psychophysiology*, vol. 28, no. 4, pp. 410–424, July 1991.

[15] M. E. Dawson, A. M. Schell, D. L. Filion, J. T. Cacioppo, L. G. Tassinary, and G. L. Berntson, "The electrodermal system," in *Handbook of Psychophysiology*. Cambridge: Cambridge University Press, 2000, pp. 200–223.

[16] M. Niepel, "Independent manipulation of stimulus change and unexpectedness dissociates indices of the orienting response," *Psychophysiology*, vol. 38, no. 1, pp. 84–91, Jan. 2001.

[17] A. Webb, *Statistical pattern recognition*, 2nd ed. West Sussex England, New Jersey: Wiley, 2002.

[18] K. P. Murphy, "Dynamic Bayesian networks: Representation, inference and learning," PhD Thesis, University of California, Berkeley, USA, 2002.

[19] P. A. Arndt and H. Colonius, "Two stages in crossmodal saccadic integration: evidence from a visual-auditory focused attention task," *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, vol. 150, no. 4, pp. 417–426, June 2003.

[20] A. Sanders, "Towards a model of stress and human performance," *Acta Psychologica*, vol. 53, no. 1, pp. 61–97, Apr. 1983.

[21] C. D. Frith and H. A. Allen, "The skin conductance orienting response as an index of attention," *Biological Psychology*, vol. 17, no. 1, pp. 27–39, Aug. 1983.

[22] F. K. Graham, "Attention: the heartbeat, the blink, and the brain," in *Attention and information processing in infants and adults: perspectives from Human and Animal research*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1992.