Bases of Empathic Animism Illusion: audio-visual perception of an object devoted to becoming perceived as a subject for HRI

Romain Magnani, Véronique Aubergé, Clarisse Bayol, Yuko Sasa

LIG, CNRS, UMR 5217, University of Grenoble Alps, France

{name.surname}@univ-grenoble-alpes.fr, Clarisse-bayol@live.fr

Abstract

Ethically, connected objects, in particular, a social robot appearance implies to know what are the effects of its design. Western societies seem to be about embracing robots' advent in the daily living, so the Human Robot Interaction field starts facing the phenomenon we call the "empathic illusion". It supposed to appear while an object is able to switch its socio-cognitive treatment by humans by becoming a subject to them. This phenomenon may change those specific objects' status within the human social sphere. This paper's aim is to start exploring the audio-visual design dimension which influence our first impressions on an object. These dimensions combinations are expected to help better understand the robots design's effects on humans.

Index Terms: social robotic, HRI, socio-affective prosody, perception test, empathic illusion, emotional induction, design

1. Introduction

Before being perceived as animated, the social robot inherit of the audio-visual (emotional, intentional, social and cultural) attributes naturally given to the things [1]. One of the challenges of HRI concerns the models allowing to handle the attachment phenomenon in human-robot interaction which is initiated by these emotional inductions [2][3][4][5]. This attachment builds progressively a relation, which can be related eventually as companionship, the role given to this object perceived as a robot/subject being dependent on the appearance, while a wide range of "companion robots" are actually spreading.

This attachment has been observed along empathic reactions for various robots appearances and different socio-cultural belongings. While humans are asked to be brutal with robots like Pleo, a stuffed animal-like robot, they are not allowing themselves to hurt it [6]. Indeed, they seem to be concerned about the robot, showing empathic reactions with physiological markers [7] as same as observed on EEG of subjects looking at an android hand being cut [8]. This empathic reaction can even bring soldiers to prepare funerals for their destroyed tool-robot [9]. Moreover, while Boston Dynamics emphasised the technical performance of their Atlas robot by hustling and pushing it, in order to challenge its abilities to recover from disturbances, they provoked a lot of indignation on social networks as the public perceived their demonstration as a bullying. Even though that robot is known not to be human or alive it still get perceived with an ability to have feelings, whereas it obviously cannot: it's an object.

A human being involved into an interaction with this kind of technology is led to having an empathic perception of someone else. The impossibility to escape from this animistic perception is what we introduced as the "empathic illusion" [10]. The Atlas anecdote perfectly illustrates this phenomenon, bringing humans to feel a form of pain for the object. The robots' appearances, regularly given by roboticists, are mostly following the human affective sphere [11]. They so can be humanoid (like Nao), "petoid" (like Aibo, Karotz), or inherent from cartoons characters and cuddly toys (like Pleo, Paro). Following up these social robots, its designs seem to focus on the emotional impact they can have on the interaction pleasantness. Roboticists seem to play down the design's effects while it constantly influence the perception guiding our interaction. However, the uncanny valley awareness [12] progressively tends to design robots by trying to pull away instinctively some chosen shapes, for more abstract ones (like Jibo, Diya One), to take them out of the affective sphere. The design's overlayered complexity due to the humanoid form as seen for Atlas, is increasing the risk of anthropomorphic projection which might affect the empathic effects on the humans' perceptions.

This paper is a perceptive inception of the design approach, for French culture, by looking at the appearances dimension, which are making our first impressions on an object that could motivate the premise of animism. Instead, to associate the appearance with emotions, subjects are asked to associate the visual stimuli to acoustic stimuli, which have been referenced in previous studies for their relation effects in HRI. We expect this perceptual rupture shall be at least partially caused by visual primitives. However, we also suppose these visual forms might induce different impressions while they are associated with acoustic stimuli which are reduced only to meaningful sounds but without the lexical content to avoid their semantic influence. These sounds are explored on acoustic impressions by keeping their socio-affective "pure prosodic" information, as they have previously been defined as possible language primitives tools to build a non-dominant and altruistic relation, which dynamics processes' results on a socio-affective glue [13]. This relation building could thus be considered as the beginning of another consideration on which the changes can be firstly influenced on the objects impression itself (with its whole characteristics: colour, size, voice, etc.)

The present study is proposed as a perception test based on the supposed basics properties of visual appearances (shape, colour, size) by associating them with sounds dynamics, by hypothesising the gain of animism through their prosody. This study is thus settled in an impressionistic approach of audio-visual combination influencing the perception of an object.

2. Perception test

The perception test consists in associating a visual object with a sound. The tested parameters and values of the study are summarised in Table 1.

2.1. Audio stimuli characteristics as references

One test is composed of 64 different sounds. These audio stimuli are pure prosodic mouth noises carrying from which the socio-affective information was reproduced by copying the prosody referenced in the French E-Wiz corpus [14], previously labelled (visually and acoustically as the originally collected stimuli were multimodal) and auto-annotated [15][16]. They were also perceptively tested on: a cultural discrimination on linguistic/control degree criteria [17], the informational values perceived intra-culturally [18] and interculturally [19], then gradually tested on the socio-affective hypothesis in HRI [11]. The selected prosodies were the most universal and resistant ones in the previous studies. The sounds were equally produced by two French speakers (male and female). Each speaker produced 4 natures of pre-lexicalised sounds ("ah", "euh", "hum", "waouh"), which prosodies are out of the emotional dominance dimension. These 4 prelexical stimuli carry negative or positive valence different labels, giving 8 distinctive sounds.

Table 1: Test parameters' values and codes

Test parameters	Values
Nature	ah, euh, hum2, waouh
Valence	positive, negative
Gender	male, female
F0	original (O), F0 modification ratio (P)
Amplitude	original, normalized (up, down)
Shape	round (R), round-sharp (N), sharp (S)
Colour	white (w), red (r)
Size	big (Bg), small (Sm)

These test stimuli were produced in an acoustically isolated room, with a portable H6 zoom microphone (mono, .wav file, 16000Hz). The recordings present natural variations of the signal's intensity and amplitude due to the recording settings. In order to control these variations, the amplitude was normalised. The normalisation consisted in amplify or reduce the signals with the Audacity tool, to reach the calculated average amplitude of all the original acoustic cues, giving 16 sounds with original and normalised amplitude.

Finally, the sounds fundamental frequency was also modified to change the voice aesthetics, which is itself a part of an object design possibly changing our perception. An algorithm is applied on the sounds' F0, to change its value, without modifying its prosodic contour, in order to keep the same socio-affective information. The female voices were augmented in pitch, and inversely for the male voice, to avoid the opposite artefacts creation (artefacts as low female voice and high pitched man voice). The modifications were pushed until a perceptive limit of a human vocal tract production. The pitched female voice ratio fluctuate between 1.08 to 1.18, while the pitched male voice fluctuate between 0.76 to 0.80. The 16 previous sounds, into these two aesthetics versions for both genders, give the 64 test sounds.

2.2. Visual stimuli characteristics

The visual cues to associate to the sounds are 12 figures. Each one is determined by three appearance parameters, combined out of 3 shapes (round; round-sharp; sharp), 2 colours (white; red), 2 sizes (big; small). In psychological approaches, moods or emotions were most of the time associated with colour, with for instance the red associated with a high arousal [20], similarly for both women and men [21]. Hence, the objects appearances dimension impressions are also coupled with other modalities as audio inputs, carrying emotional semantic. Thus the chosen red (RGB code: R: 255 G: 0 B: 0 / HSV: 0° 100% 100%) is a very intense one with very high saturation and high value/brightness. It is opposed to white (RGB code: R: 255 G: 255 B: 255 / HSV: 0° 0% 100%) which appears by contrast on a neutralized grey background (RGB code: R: 189 G: 189 B: 189 / HSV: 0° 0% 75%).

The three types of shapes – round, sharp, and "round-sharp" (an intermediary blending the two others) – appear in an impressionist paradigm. The choice of the sharp shapes is motivated by the wide observations of the kiki/bouba experiment [22]. This study has brought to light an ideasthesia effect [23][24][25] between shapes and words. This kind universal sounds impressionism were also depicted for colours [26] with relatively resistant phonemes/colours association, even synaesthesia in music [27], or textures associated with voice quality [28]. But as the colours perception have been shown to mostly depend on socio-cultural backgrounds discussed as in [29] based on the Sapir and Whorf language relativity [30] this dimension is expected to have potential side coupled effects with the sounds gender for instance. In the present study, each shape has a red and a white prototype. The red spiky shape is motivated to be more often matched with negative sounds, and white round shapes with positive sounds which are most culture-independently shared impressions [22]. This might also motivate the design of the animation characters or the "companion robots", which tend to be rounded and white. Likewise, smaller things are expected to emit a higher pitched sound due to the vocal track congruence [31]. Little shapes are so expected to be more often matched with high-pitched sounds, and big shapes with low-pitched sounds.

Complementarily, a short qualitative study has been done in parallel to this perception test, in order to determine how humans qualify the 12 figures without the audio stimuli. The 11 participants (not participating in the perception test) firstly describe each figure with their own words. Secondly, they did an association task on the following propositions: kindness, aggressiveness, valency, potency, dominance, aesthetics and gender. As results, the participants tend to perceive big shapes as more protective, but also more aggressive and dominant than little ones. Red shapes had similar effects over the white ones. Globally, rounded shapes seem to be more likely inoffensive. Only among the little shapes, rounds were sometimes perceived bigger than round-sharps (this effect vanished for the big ones).

3. Experimental settings

3.1. Stimuli presentation

The 64 sounds were presented one by one to be associated each time with one shape among the 12 visual figures presented all at once on a web browser set as in Figure 1. This same interface is kept during a test (no figures position's variation), but the figures configuration and the sounds both changed randomly between judges. There is also a control subset of judges who used one specific set of interface/sound order, which is used to verify the eventual learning effects.

For each stimulus, the judge: 1) click on a "play" button to hear the sound (which can repeatedly played until the validation of the associated figure), 2) choose the best suiting figure to the sound (the figure can be changed until its validation), 3) put a grade on the answer's confidence level using a Likert scale (from 1: lowest - to 5: highest). Then the interface switched to the next sound. There are no restrictions on the judges' socio-cultural origins criteria.

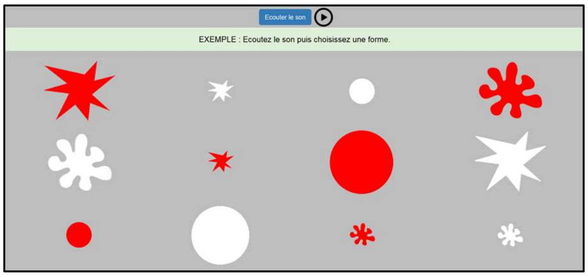

Figure 1: *Screenshot from one of the perception test interface's possible layout.*

The traces collected are: the socio-cultural information; the test start and end times; the response reaction times; the interface actions traces; the chosen values on each stimulus; the interface settings code the interface configuration and sound randomisations; the chosen figure/sound couple confidence degrees.

4. Results

4.1. Collected data samples

A total of 93 judges composed the test panel, representing 5952 pairings. They are mostly French, divided into 62% women and 38% men, with 48% within 25 and 35 years old. The other 52% are within 6 to 75 years old. There are as many participants who are used to manipulate audio-visual material in their job or spare time as the one who does not. Concerning the other cultures participants, the test was completed by 17 subjects within 9 foreign cultures and 3 biculture pairs. As the data samples present a wide variation of cultures but a few participants in each of them, the analysis will focus only on the 77 French subjects. The dataset corresponds to 2668 pairings of 35 subjects in the unrandomised subset and 2240 pairings of 42 subjects in the randomised subset. The subset is firstly analysed separately and is then merged to control the learning effects on the unrandomised subset samples.

4.2. Congruence vs. Incongruence phenomenon

We observed some strong attractions between variables and contrastively several discrepancies. Indeed, it appears that some sounds have global incompatibilities (i.e a specific shape including its colour and size), and some others have specific incompatibilities (i.e. only a colour, or only a size, or only a colour and size, etc.) In contrast, we see global and specific compatibilities. These results could be summarised as follows:

- Idiosyncratic Congruence / Thematic Congruence
- Idiosyncratic Incongruence / Thematic Incongruence

Table 2 illustrates those strong attraction/repulsion phenomena. Each cell presents the association percentage and the codename (cf. Table 1) for the shape appearance.

Table 2: *Strong and weak associations of sounds and appearances for the 77 French judges group*

Normalised amplitude only (to minimize the recording condition effects)		**Shape_Color_Size**			
		Positive		**Negative**	
		Strongest association	**Weakest association**	**Strongest association**	**Weakest association**
Female-P	**ah**	36% / S_r_Bg	0% / *_w_Sm	19% / R_w_Bg	1% S_r_Bg
	euh	29% / N_w_Sm	1% / S_w_Bg	25% / R_w_Sm	0% *_r_Bg
	hum2	55% / N_w_Sm	0-2% / *_r_*	31% / N_r_Sm	0% / R_*_Bg
	waouh	33% / S_w_Bg	~2% / *_*_Sm	22% / R_w_Bg	3% / S_r_Bg
Female-O	**ah**	27% / S_r_Bg	~3% / *_*_Sm	23% / R_w_*	1% / S_r_Bg
	euh	~23%/ N_w_*	~3% / *_r_*	21% / R_w_Sm	~2% / S_*_Bg
	hum2	42% / N_w_Sm	~3% / *_r_*	~25% / S_*_Sm	~3% / *_*_Bg
	waouh	30% / S_w_Bg	~2% / *_*_Sm	~20% / R_w_*	~4% / *_r_*
Male-P	**ah**	21% / N_w_Bg	1% / S_r_*	23% / R_r_Sm	~1% / S_*_Bg
	euh	25%/ N_r_Bg	~2% / S_*_*	~15% / *_*_Sm	~3% / *_*_Bg
	hum2	25% / N_r_Sm	~1% / S_*_Bg	25% / R_r_Sm	~2%/ *_*_Bg
	waouh	34%/ S_r_Bg	~2% / N_*_*	23% / R_r_Bg	~4% / N_*_Sm
Male-O	**ah**	23% / N_w_Bg	~3% / *_*_Sm	19% / R_r_Sm	0% / S_*_Bg
	euh	~19% / N_*_Bg	1% / S_w_Sm	~18% / R_*_Sm	~1% / S_*_Bg
	hum2	22% / N_*_Bg	~2% / S_*_*	25% / S_r_Sm	~1% / R_*_Bg
	waouh	~25% / S_*_Bg	~2% / *_*_Sm	17% / R_w_Sm	0% / N_r_Bg

An asterisk means that one of the appearance dimensions (described as shape_color_size) is not significant. This table also depicts that congruencies are more often idiosyncratic, while incongruences are more often thematic. For example, the positive Female-P hum2 have a solid idiosyncratic congruence with the little white round-sharp (N_w_Sm), and an also strong thematic incongruence with red appearances (*_r_*). One of the present perception test motivation was to explore the covariance/variance of the audio and visual stimuli parameters. As observed in this contingences matrix, this methodology does not show an exact correspondence between one acoustic parameter and one visual parameter as the dynamics of combination could affect the perception differently. In the next section, we applied Manovas in the R program in order to observe which acoustics dimensions coupling might influence the appearances' choice.

4.3. MANOVA applications

In order to explore this potential combination effects, MANOVAs were applied with sounds as factors, in uncombined condition (on 5 acoustics dimensions: nature, valency, gender, amplitude, voice change with F0) then exhaustively combined into 2 to 5 dimensions (corresponding to 27 sounds vectors). Each of these sounds vectors are crossed with the 3 appearance variables (shape, colour, size), also firstly uncombined then combined in pairs (shape-colour, shape-size and colour-size) then both of three altogether. This leads to 198 vectors tested for each of the three subsets' groups (the randomised, the unrandomized and the 77 French subjects group). All the MANOVAs are tested on the 4 Hotteling's-Lawley, Pillai, Roy and Wilks tests and are also analysed by decomposition. Moreover, the four applicability conditions (including independence and normality) were previously verified.

4.3.1. Uncombined acoustic dimensions behaviors

At first, the MANOVAs seem to be robust to analyse the decomposed dimension effects, but only until two combined dimensions, as significant differences between are noticed. But over two dimensions, the effects of each parameter are too strong to see the changes and so all combinations appeared to be significantly linked to the figure choice. Globally, the randomised group and the whole French group are showing the same characteristics. The unrandomised interface group shows the same tendency, but with additional combinations minimising the effects of certain choice. By looking at the parameter through the MANOVAs decomposition, each sound variable (nature, F0, gender, valence and amplitude) presents privileged or avoiding appearance parameters.

Concerning the sounds' nature, there are no noticeable differences, as it always appears as significant in the figure choice ($p<0.001$). We so suppose the information values carried in this parameter to be too rich to isolate focused effects on the forms' choices.

One of the most varying parameters is the F0 giving the voice aesthetics changes. The colour choice is preferentially associated with this F0 variation ($p<0.05$) with a stronger effect if it is associated with a size variation. However, it is the combination (and not only the size) which can explain this choice. This observation concerns all the groups, including the total participants, adding the foreigners. The gender seems to be never significant on the size choice while considering only this dimension. Moreover, the size is neither non-significant for the unrandomised group, only the colour explains the choice from gender ($p<0.001$). Besides, gender and valence have inversed behaviours. While the valence is combined with the two others parameters, the colours are explained significantly the choice ($p<0.001$). Besides, the amplitude seems to be not significant for the shape choice or its association with colours. It is even only justifying the size in the control group. Its effects seem to be minimised while combined, which we can consider is due to the little changes of the amplitude normalisation. Finally, while a choice is highly significant on one appearance basing on one acoustic parameter, its effects seem to appear strongly to change the other parameters' behaviours.

4.3.2. Acoustic dimensions combination effects

While we consider the effects of two sounds parameters combination on the choice of uncombined and combined appearance parameters, the major visible effects concern the pairings with the F0. While this dimension is combined with gender, it is not avoiding the choice of the size alone, and by side effects as seen before, a less significant association with combined or not shape. This effect is amplified for the unrandomised interface group. Moreover, as the effect is only minimised for the control group, it is strengthened for the group with all participants. The amplitude combined with F0 has a minimised effect on the shape and its combination with colours, and even for the shape/size pair for the randomised group (more than 0.05 significant p-values). For the control group, this effect is confirmed with a significant choice only visible on the size. Finally, the valence combined with F0 avoids mostly colours for all groups.

5. Conclusions

In this study, we initiate a research perspective to understand what causes an object to become perceived as a subject, by looking first for the primitives of multimodal appearances which are making our first impressions on an object by modifying its socio-cognitive treatment. The data analysis of the perception test illustrated a congruence vs. incongruence phenomenon between sounds and static shapes: with some pairs strongly chosen by the panel and some others systematically dodged. This could be the first step to enlighten the empathic illusion phenomenon, which is introduced to occur while the object is modifying its socio-cognitive state regarding humans observing it. As one can hypothesise, a consistency of various ontologies including movement (in this case it's sound/shape but it could be other dimensions) is mandatory to induce animism. In a second part, MANOVAs showed some strong effects on some acoustic parameters as F0 for instance. However, the combination between two acoustic features seem to have more complex interactions and could be completed by other approaches as they might not follow a statistical linear regression law as proposed in the MANOVAs. We assume that the social robotic field needs to develop solid knowledge about these emotional inductions to better understand the attachment phenomenon in human-robot interaction and to consider it better for the sake of an essential ethical approach.

6. Acknowledgements

We would like to thank Nicolas Bonnefond (Amiqual4Home, Inria) for his technical assistance, and Ambre Davat (Unniversity of Grenoble Alps, PhD student) for sharing her sounds conversion algorithm.

7. References

[1] N. H. Frijda, « Emotion, cognitive structure, and action tendency », *Cogn. Emot.*, vol. 1, nº 2, p. 115–143, 1987.

[2] S. M. Anzalone, S. Boucenna, S. Ivaldi, and M. Chetouani, « Evaluating the Engagement with Social Robots », *Int. J. Soc. Robot.*, p. 1–14, 2015.

[3] J. Hall, T. Tritton, A. Rowe, A. Pipe, C. Melhuish, and U. Leonards, « Perception of own and robot engagement in human–robot interactions and their dependence on robotics knowledge », *Robot. Auton. Syst.*, vol. 62, nº 3, p. 392–399, 2014.

[4] C. Rich, B. Ponsler, A. Holroyd, et C. L. Sidner, « Recognizing engagement in human-robot interaction », in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference*, p. 375–382, 2010.

[5] C. L. Sidner, C. D. Kidd, C. Lee, and N. Lesh, « Where to look: a study of human-robot engagement », in *Proceedings of the 9th international conference on Intelligent user interfaces*, p. 78–84, 2004.

[6] K. Darling, « Extending Legal Rights to Social Robots », Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2044797, 2012.

[7] A. M. Rosenthal-von der Pütten, N. C. Krämer, L. Hoffmann, S. Sobieraj, and S. C. Eimler, « An Experimental Study on Emotional Reactions Towards a Robot », *Int. J. Soc. Robot.*, vol. 5, nº 1, p. 17-34, 2012.

[8] Y. Suzuki, L. Galli, A. Ikeda, S. Itakura, and M. Kitazaki, « Measuring empathy for human and robot hand pain using electroencephalography », *Sci. Rep.*, vol. 5, 15924, 2015.

[9] J. Carpenter, « Culture and Human-Robot Interaction in Militarized Spaces: A War Story », *Routledge*, 2016.

[10] V. Aubergé, « Illusion empathique du robot social : réparation de "l'autre" fantôme », *7e rencontres du Pôle Grenoble Cognition*, Grenoble, France, 2016.

[11] V. Aubergé *et al.*, « The EEE corpus: socio-affective "glue" cues in elderly-robot interactions in a Smart Home with the Emoz platform », *5th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data*, Reykjavik, Iceland, 2014.

[12] M. Mori, « La vallée de l'étrange », *Gradhiva*, nº 1, p. 26–33, 2012.

[13] Y. Sasa and V. Aubergé, « Socio-affective interactions between a companion robot and elderly in a Smart Home context: prosody as the main vector of the "socio-affective glue" », *Speech Prosody 7*, Dublin, Ireland, 2014.

[14] V. Aubergé, N. Audibert, and A. Rilliard, « E-Wiz: A trapper protocol for hunting the expressive speech corpora in lab », *4th International Conference on Language Resources and Evaluation*, Lisbonne, Portugal, p. 179-182, 2004.

[15] F. Loyau and V. Aubergé, « Expressions outside the talk turn: ethograms of the feeling of thinking », in *5th LREC*, p. 47–50, 2006.

[16] A. Vanpé and V. Aubergé, « Early meaning before the phonemes concatenation? Prosodic cues for Feeling of Thinking », *GSCP Belo Horiz. Bresil*, 2012.

[17] R. Signorello, V. Aubergé, A. Vanpé, L. Granjon, and N. Audibert, « À la recherche d'indices de culture et/ou de langue dans les micro-événements audio-visuels de l'interaction face à face », in *WACA 2010*, p. 69–76, 2010.

[18] G. D. Biasi, V. Auberge, and L. Granjon, « Perception of social affects from non-lexical sounds », in *GSCP*, 2012.

[19] Y. Sasa, V. Aubergé, and A. Rilliard, « Social micro-expressions within Japanese-French contrast », *WACAI*, Grenoble, France, 2013.

[20] R. Plutchik, « The Nature of Emotions », *Am. Sci.*, vol. 89, nº 4, p. 344, 2001.

[21] A. Mehrabian, « Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament », *Curr. Psychol.*, vol. 14, nº 4, p. 261-292, 1996.

[22] W. Köhler, *Gestalt psychology*. New York: H. Liveright, 1929.

[23] E. Milan, O. Iborra, M.J. de Cordoba, V. Juarez-Ramos, M. A. R. Artacho, and J. L. Rubio, « The Kiki-Bouba Effect A Case of Personification and Ideaesthesia », *J. Conscious. Stud.*, vol. 20, nº 1-2, p. 84-102, 2013.

[24] D. Nikolić, « Is synaesthesia actually ideaestesia? An inquiry into the nature of the phenomenon », in *Proceedings of the Third International Congress on Synaesthesia, Science & Art*, p. 26–29, 2009.

[25] A. Mroczko-Wąsowicz and D. Nikolić, « Semantic mechanisms may be responsible for developing synesthesia », *Front. Hum. Neurosci.* vol. 8, 2014.

[26] K. Watanabe, Y. Greenberg, and Y. Sagisaka, « Sentiment analysis of color attributes derived from vowel sound impression for multimodal expression », in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, 2014, p. 1-5.

[27] J. Ward, B. Huckstep, and E. Tsakanikos, « Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? », *Cortex J. Devoted Study Nerv. Syst. Behav.*, vol. 42, nº 2, p. 264-280, February, 2006.

[28] A. Moos, D. Simmons, J. Simner, and R. Smith, « Color and texture associations in voice-induced synesthesia », *Front. Psychol.*, vol. 4, p. 568, 2013.

[29] P. Kay and T. Regier, « Language, thought and color: recent developments », *Trends Cogn. Sci.*, vol. 10, nº 2, p. 51-54, February, 2006.

[30] P. Kay and W. Kempton, « What is the Sapir-Whorf hypothesis? », *Am. Anthropol.*, vol. 86, nº 1, p. 65–79, 1984.

[31] J. J. Ohala, *The frequency code underlies the sound symbolic use of voice pitch. Sound symbolism, ed. by Leanne Hinton, Johanna Nichols, and John J. Ohala, 325–47*. Cambridge: Cambridge University Press, 1994.