SUMA PSICOLÓGICA

http://sumapsicologica.konradlorenz.edu.co

Ocular fixations modulate audiovisual semantic congruency when standing in an upright position

Guillermo Rodríguez-Martínez [a,*], Henry Castillo-Parra [b], Pedro J. Rosa [c,d], Fernando Marroquín-Ciendúa [a]

[a] *School of Advertising, Universidad Jorge Tadeo Lozano, Bogotá, Colombia*
[b] *Faculty of Psychology, Universidad de San Buenaventura, Medellín, Colombia*
[c] *Lusófona University, Digital Human-Environment Interaction Lab (HEI-lab), Lisbon, Portugal*
[d] *ISCTE-Instituto Universitário de Lisboa, CIS-IUL, Lisbon, Portugal*

Received 2 June 2020; accepted 16 October 2020

KEYWORDS
Crossmodal semantic congruency,
ocular fixations,
bistable perception,
body orientation

Abstract Introduction: Multisensory audiovisual semantic congruency is the process by which visual information is perceived as integrated to auditory stimuli, because both coincide in terms of simultaneity and semantic correspondence. This study was aimed at establishing whether visual percepts, which semantically correspond to auditory stimuli, are associated with ocular fixations in modulating bottom-up areas while keeping a body posture alignment between the up-direction and the idiotropic axes, as well as in another orientation corresponding to a vectorial opposition between the up-direction and the head idiotropic axis. **Method:** Two groups (one for each position) were selected from a sample of 88 people. A bistable image was presented on a screen of a fixed 120 Hz eye-tracker device, providing background auditory stimuli so as to establish semantic congruencies and their relations to ocular fixations. **Results:** It was found that audiovisual semantic congruency is associated with fixations when idiotropic vectors are aligned with the up direction. Fixations manifested in bottom-up modulating areas are not associated with multisensory audiovisual semantic congruency when the head idiotropic vector is parallel with the gravity vector. Eye fixations decrease significantly if the head idiotropic axis is aligned with the gravity vector. **Conclusion:** It is concluded that body position can affect visual perceptual processes involved in the occurrence of semantic congruency.

* Corresponding author.
E-mail: guillermo.rodriguez@utadeo.edu.co

https://doi.org/10.14349/sumapsi.2021.v28.n1.6
ISSN 0121-4381, ISSN-E 2145-9797/

Las fijaciones oculares modulan la congruencia semántica audiovisual cuando el cuerpo está en una posición erguida

PALABRAS CLAVE
Congruencia semántica, fijaciones oculares, percepción biestable, orientación del cuerpo

Resumen **Introducción:** La congruencia semántica audiovisual es el proceso por el cual información de una modalidad sensorial visual se percibe como integrada a estímulos auditivos, porque coinciden en términos de simultaneidad y correspondencia semántica. Este estudio tuvo por objeto establecer si los perceptos visuales que se corresponden semánticamente a estímulos auditivos están asociados con las fijaciones oculares realizadas en áreas de modulación *bottom-up*, tanto en una postura corporal definida por la alineación entre la dirección vertical hacia arriba y los ejes vectoriales idiotrópicos, como en otra orientación definida por una oposición vectorial entre la vertical hacia arriba y el eje idiotrópico de la cabeza. **Método:** Dos grupos fueron seleccionados (uno por cada posición), tomados de una muestra de 88 personas. Los datos se obtuvieron por medio de un dispositivo fijo de registro de movimientos oculares de 120 Hz. Una imagen biestable se presentó, proporcionando estímulos auditivos de fondo para producir congruencias semánticas y establecer su relación con las fijaciones oculares. **Resultados:** Se encontró que la congruencia semántica audiovisual está asociada con áreas de fijación ocular cuando los vectores idiotrópicos están alineados con la dirección vertical ascendente. Las fijaciones oculares en áreas de modulación *bottom-up* no están asociadas con la congruencia semántica audiovisual cuando el vector idiotrópico de la cabeza está alineado con el vector gravitacional. En esta última posición, la cantidad de fijaciones oculares es significativamente menor. **Conclusión:** La posición del cuerpo puede afectar procesos perceptuales visuales que, a su vez, inciden en el efecto de congruencia semántica.

The phenomenon of multisensory audiovisual semantic integration is a process through which information of a sensorial visual modality is perceived as integrated with stimulus of an auditory nature due to the similarity in terms of semantic load (Hsiao et al., 2012; Smith et al., 2007). The audiovisual semantic congruency effect is normally assessed by measuring outputs observed when presenting matching and mismatching images and sounds (Spence, 2007, 2011). For instance, bistable images have been used so as to assess the semantic congruency effect (e.g., Hsiao et al., 2012; Smith et al., 2007), given the fact that these images can be interpreted in two different ways in terms of semantic content (Rodríguez-Martínez & Castillo-Parra, 2018a). The correspondance is observed when it can be assumed that both, the auditory and the visual stimulus, constitute a perceptual unity, also called unity assumption (Vatakis & Spence, 2008). It is possible to use semantic context by stimulating auditorily through soundtracks while observing bistable figures for a long period of time (Smith et al., 2007). That context can be provided by tones of voice, as long as the content of the message itself is unclear because of the unfamiliarity with the language that is being used (Hsiao et al., 2012). The influence exerted by the tones of voice can operate as a top-down modulating factor (Rodríguez-Martínez & Castillo-Parra, 2018b), despite the fact that there can be bottom-up modulating features that affect the interpretation of the visual stimulus like ocular fixations in areas that favour its perception (Hsiao et al., 2012). In this regard, Gale and Findlay (1983) demonstrated that there are critical areas in the bistable image *my girlfriend or my mother-in-law* that influence the perception of its two possible percepts. It was suggested that certain areas enable the observer to perceive one image more than the other (young woman or old woman), as can be seen in Figure 1. Thus, the A1 area, which modulates the young woman percept, contains defining lines of the eye and nose of the young woman; the A2 area predominantly defines the ear of the young woman and the eye of the elderly one; the A3 area, modulating the percept of the elderly woman, refers to the elderly woman's mouth; the A4 area contains a line that defines the elderly woman's nose and, at the same time, a contour of the young woman's chin. Fixating on the lines and features of a bistable image can influence its final perception (Brouwer & van Ee, 2006).

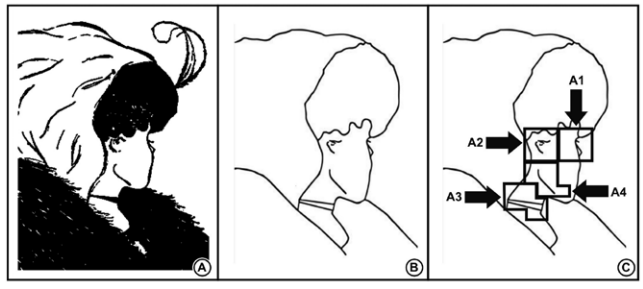

Figure 1. The bistable image *my girlfriend or my mother-in-law*. (A) The original version of the image. (B) The version that was used by Gale and Findlay (1983) so as to determine bottom-up modulating areas. (C) Bottom-up modulating areas. Sources: Image A was adapted from García-Pérez (1989). Images B and C were adapted from Gale and Findlay (1983).

Bistable images have also contributed to understanding the phenomenon of semantic congruence (e.g., Hsiao et al., 2012; Yeh et al., 2011). This has been done by observing correspondences or non-correspondences between auditory semantic content and visual percepts identified from the bistable image used (e.g., Marroquín-Ciendúa et al., 2020). Likewise, bistable visual stimuli have been useful in order to unravel the modulating role that body spatial orientation might play in perceptual processes (e.g., Clément & Eckardt, 2005; Yamamoto & Yamamoto, 2006). The existence of an

impairment in eye movements has been suggested due to variations in orientations of both body and head (Clarke, 2008). On the other hand, multimodal tactile, proprioceptive, and visual mechanisms have been found in relation to perception of bistable images (Yamamoto & Yamamoto, 2006). Consequently, bistable visual processing involved in an audiovisual perceptual task can condition the possibility that the effect of semantic congruence occurs (Hsiao et al., 2012). This might happen if, by affecting the position of the body, visual interpretations of bistable images are modulated while occurring congruence or incongruence with a given auditory stimulation. Observing this modulating effect of body position on visual processing associated with an audiovisual multisensory task could contribute to the understanding of the effect that body posture may have on audiovisual perceptual integration.

Body spatial orientation and its effects on perceptual processes

The gravity vector (G vector) follows the trajectory to the fall of a solid object in conditions of 1-g force (Mittelstaedt, 1983). There is an opposite direction relative to this vector, the so-called 'vertical' (Lopez et al., 2007). The point that defines the destination of the vertical vector is called physical zenith (PZ) (Mittelstaedt, 1983). Besides, there is a subjective visual vertical, which is the perceptual visual vertical (also perceptual upright), normally determined by visual cues, and also by the idiotropic tendency to perceive the visual vertical aligned with the body axis in a footward direction (Haji-Khamneh & Harris, 2010; Oman, 2003). The vector that points to the physical zenith (the up direction itself) is not always aligned with the subjective visual vertical, also called subjective zenith (SZ) (Mittelstaedt, 1983). When referring to the idiotropic vector (also known as Z body axis), what is implied is the direction of the components of the body considered as a whole (Lopez et al., 2008). Two different idiotropic axes have been recognized: the idiotropic axis of the trunk (Zt) and the one that refers to the head (Zh). These two axes are aligned when a person is perfectly upright; however, when the head is down, or when looking up while standing, the head idiotropic axis points in a different direction relative to the idiotropic axis of the trunk (Mittelstaedt, 1983). In brief, different vectorial magnitudes have been established, as can be seen in Figure 2.

These vectors have been used in order to have reference points that allow defining body positions for studies concerning vestibular inputs (e.g., Oman, 2003), spatial perception (e.g., Lopez et al., 2008), effects of body posture on perception (e.g., Yamamoto & Yamamoto, 2006), among others.

As far as vestibular inputs are concerned, it has been stated that they involve complex brain functions through interactions with inputs concerning other senses (Ferrè et al., 2015). Vestibular-visual interactions have been understood to be an intermodal combination of cues, while underlying perceptual dimension of heading direction (Yamamoto & Yamamoto, 2006). What is implied here is a multisensory modulation because of the existence of a modulating process through which a sensory signal exerts an influence on a second sensory pathway (Ferrè et al., 2015). Vestibular-visual interactions and specific circuits associated with visual perception are activated when the perception of equilibrium and spatial position involves corrections or adjustments

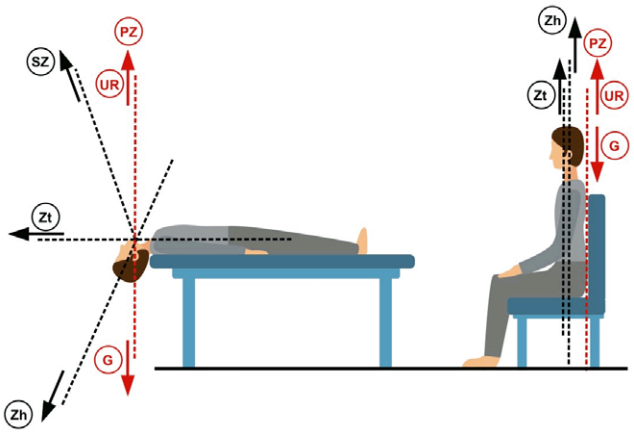

Figure 2. Vector directions involved in body orientation. It can be seen how, in an up-right posture (on the right), the idiotropic vectors (Zh and Zt) are aligned with the vector traced towards physical zenith (PZ); that is to say, with the up-right direction (UR). On the left, six vectors are illustrated as follows: 1. Idiotropic trunk vector (Zt); 2. Up-right direction (UR); 3. Physical zenith (PZ); 4. Gravity vector (G); 5. Subjective zenith (SZ); 6. Physical zenith (PZ). Source: Own design adapted from Mittelstaedt (1983).

concerning eye-movements (Angelaki et al., 2009). Besides, a proximal stimulus significantly varies in relation to body posture; this fact involves perceptual processes while viewing a bistable image (Clément & Eckardt, 2005; Yamamoto & Yamamoto, 2006). As the idiotropic head axis is inverted, the eyes' vertical meridians are rotated with regard to the gravity vector, which implies an effect on visual perception (Gaunet & Berthoz, 2000). Given that unusual body positions imply an effect on the way of interacting with the perceived environment (Balint & Hall, 2015), a research question was raised of whether body position exerts an influence on audiovisual semantic congruency, taking into account ocular fixations made on bottom-up modulating visual areas.

Analyses were made in order to determine if when a percept is reported as semantically congruent with a modulating audio, the observed area of fixation is the bottom-up modulating area that favours that percept, considering two different body postures: 1. A position where the head idiotropic axis (Zh) is parallel to the up-direction; 2. A position where that idiotropic axis is aligned with the gravity vector. It was hypothesized that the areas on which eyes are fixated are associated with the visual percept that is recognized, but only with regard to the body position in which the vertical up direction and Zh axe are aligned. Another hypothesis was made in the sense that ocular fixation areas modulate the recognition of the visual percept that is congruent with the semantic load of auditory stimulation. Finally, another hypothesis was proposed in the sense that more ocular fixations occur in the position in which the idiotropic axes point towards the physical zenith.

Method

Participants

The sample consisted of 88 undergraduate students who were studying at universities in Medellín, Colombia (average

age, *M* = 20.98, *SD* = 1.85; 58% men; 42% women). All of them reported not having had medical histories involving cerebral and/or vestibular system damage, panic and vertigo disorders, or hypertension or hearing problems. Moreover, their self-reported medical record implied not having been diagnosed with visual problems or with the need to use devices to correct visual impairments. Each participant gave their informed consent prior to the experiment, and received monetary payment upon its completion. The study was approved by the ethical committee of San Buenaventura University in Medellin.

Participants were divided in two groups (see Figure 3): the first one had to carry out the task in a body position defined by the alignment between the up-right direction (UR) and Zh axis (Z axis projected from head axis). This position was named non-inverted position (NIP). The second group performed the task in a body position that corresponded to a vectorial opposition between the up-right direction and Zh axis, where the Zt axis was at a rotation angle of 20^0 relative to the horizontal. This position was called inverted position (IP).

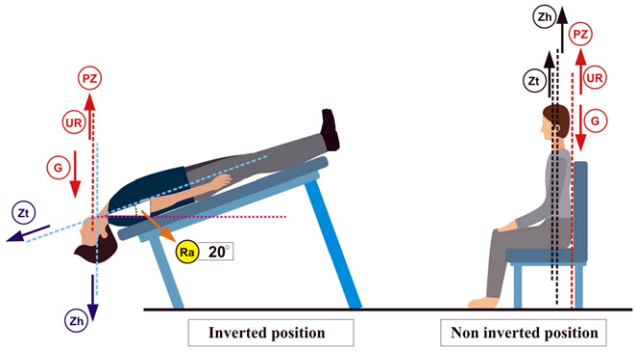

Figure 3. The two body positions considered. On the left, the position related to the IP group (inverted position). The rotation angle (Ra) of the idiotropic trunk axis (Zt) in relation to horizontal line was a 20-degree angle. On the right, the body posture of the NIP group (non-inverted position).

Apparatus and stimuli

A fixed 120 Hz eye-tracker device, reference standard Tobii™ T120, was used to record the data. *Creative HN-900* headphones were used in order to deliver the sounds that were selected as auditory modulating stimulation. The bistable image *my girlfriend or my mother-in-law* adapted by Gale and Findlay (1983) was the visual stimulus used. There were two auditory stimuli: the voice of a young lady pronouncing words in French, at 52 dB SPL, and the voice of an old woman, also speaking in French, at 52 dB SPL. Using Avid Pro Tools HD 12.6, the two audios were compressed to achieve a balanced level in the waves. The audios were validated previously: both voices were listened to by 161 people (average age, *M* = 21.01, *SD* = 2.36; 67.7% women; 32.3% men) who had similar sociodemographic conditions to the participants who took part in the experiment. The average age given to the person who emitted the voice of the young lady was 23.83 (*SD* = 4.93), while the average age given to the voice of the old woman was 82.7 (*SD* = 8.12*)*; *t* (320) = 78.66*;* $p < .001$). The audiovisual integration effect that we wanted to observe was based on tones of voice rather than the semantic charge of the spoken words (at the end of the test, as a control, participants were asked if they understood any of what had been expressed by the voices, so as to ensure that only the tones of voice operated as modulating auditory stimulation). All of them complied with this requirement.

Procedure

It was necessary to position the participants (both groups) so that their faces were parallel to the monitor of the eye-tracker device. A viewing distance of 60 cm was the measurement deemed appropriate for the calibration phase, for both groups. All of the participants viewed the bistable image while listening to the voices separately. The experimental test was done in such a way that the stimulus on the retina (retinal image) was the same for both groups (this means that the visual stimuli for the IP group were inverted on the monitor, as shown in Figure 4). Each participant had to report (by clicking a mouse button) when they began to perceive one or another of the possible percepts. Participants had to continuously report the visual percepts identified, saying the words “young” or “old”, as appropriate. These reports were requested each time they began a perceptual recognition. The image was presented to each participant twice for 20 seconds each time. During one of the exposures, the audio of the young lady’s voice was heard. During the second exposure of the image, the elderly female’s voice was heard. The order in which the audios were delivered was counterbalanced. There was a 4 second interval between each exposure to the audiovisual stimulus. Prior to initiating each audiovisual stimulus, a fixation point was presented on the screen for 200 milliseconds. This fixation point did not favour the perception of either of the percepts (e.g., Hsiao et al., 2012). Figure 4 illustrates this procedure.

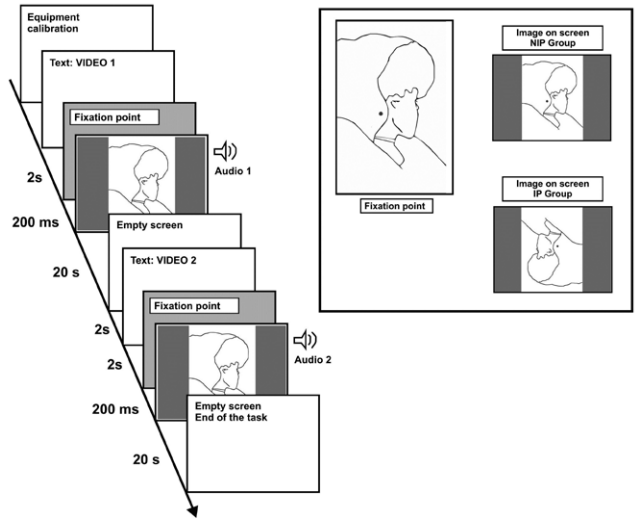

Figure 4. Procedure. The left-hand side shows the succession of the complete procedure, from the calibration phase to the end of the task. The right-hand side shows the location of the fixation point and the images on the screen for each group.

The areas of interest (AOIs) were codified as A1, A2, A3, A4, and "B" (Background), as can be seen in Figure 5. In order to assess audiovisual integration, it was necessary to count the number of visual percepts that were semantically congruent with each tone of voice. Based on the records, the perceptual reports were shown in a data table for each participant, indicating the area of interest on which their gaze was fixed 250 milliseconds before the recording of each report. This subtraction was made considering the time-difference that existed between each report and the moment in which the perceptual recognition was made, which involved a complex reaction time (Bonnet, 1994). In order to establish relations between fixation areas and congruences and incongruences with the audio involved, analyses of association were carried out between each of the AOIs and the congruent and incongruent visual percepts relative to each tone of voice used. All analyses were conducted using SPSS software (v.23) for Windows.

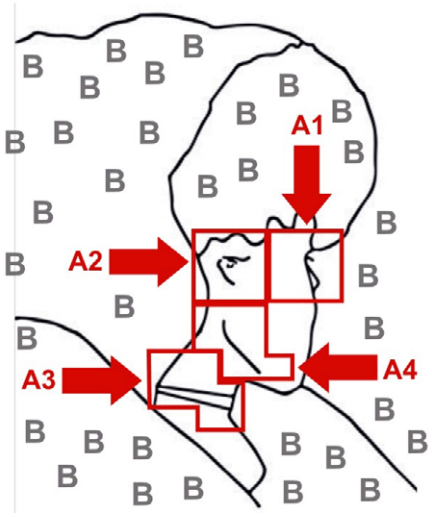

Figure 5. The areas of interest defined to make the analyses.

To define the sample size of ocular fixations, what is stated in Cohen (1992) was taken into account regarding the sample size that is necessary to achieve a small (.10) medium (.30) or large (.50) effect, with power of .80, for chi-square tests. Thus, in order to obtain a power of .80 and a medium effect of .30, a sample of ocular fixations greater than 130 was proposed. Based on chi-square analysis (χ^2 = 18,617 (4), *Sig* = .001, *1-β* = .81, *W* = .33), it was observed that there was a significant difference with a statistical power that exceeded the standard of .80. What's more, the size of the effect exceeded the average standard of .30. Likewise, the contingency and independence coefficient yielded a value of .274, and a significance of .001. As far as the statistical power and size effect for *t student* tests are concerned (ocular fixations comparisons), it was estimated to achieve a statistical power of .80, and a medium size effect of .50. The power value was lower than .80 (*1-β* = .72; *d* = .48).

Results

Analysis of visual percepts congruent with the audio and associations with ocular fixation areas

Results for group NIP

The semantic congruence that occurred between audios and visual percepts were associated (χ2 (12, *N* = 214) = 26.689, *p* = .009) with ocular fixations made in areas of interest (see Table 1). With respect to the A1 area, the congruence concerning the young woman's voice had greater association with ocular fixations, being statistically significant (*p* < .05) in comparison with the incongruence regarding that auditory stimulus. 37% of the congruence found between the audio and the young woman (YW) percept occurred when the A1 area was observed (see Tables 1 and 2).

The ranked associations test (Table 2) shows how the fixations in area A1 that were congruent with the YW percept were significantly different (*p* < .05) in comparison to when they corresponded to the incongruent percept. On the other hand, the association between the incongruence for the OW percept when observing the A1 area was significantly different when compared with the congruence for the OW percept, as well as when it was compared with the incongruence with YW (regarding ocular fixations in the A1 area).

Table 1 **Semantic congruence for each reported percept considering the AOIs (NIP group)**

AOIs	YW Cong.	% YW Cong.	YW Incong.	% YW Incong.	OW Cong.	% OW Cong.	OW Incong.	% OW Incong.
A1	23	37%	5	12%	7	15%	23	38%
A2	18	29%	22	51%	24	50%	26	43%
A3	4	6%	5	12%	5	10%	1	2%
A4	5	8%	4	9%	8	17%	3	5%
Background	12	19%	7	16%	4	8%	8	13%
Total	62	59%	43	41%	48	44%	61	56%

Note: Test with Chi-squared statistic (χ^2 (12, *N* = 214) = 26.689, *p* = .009). Cong = Congruent; Incong = Incongruent. AOIs= Areas of interest. YW= Young woman percept. OW= Old woman percept.

Table 2 Results association tests between fixations in areas of interest and congruences (audiovisual integrations) and incongruences (NIP group)

AOIs	Congruence with YW (A)	Incongruence with YW (B)	Congruence with OW (C)	Incongruence with OW (D)
AOIs A1	B*	-	-	B*C*
AOIs A2	-	-	-	-
AOIs A3	-	-	-	-
AOIs A4	-	-	-	-
Background	-	-	-	-

Note: Test with statistics from association tests. AOIs= Areas of interest. YW= Young woman percept. OW= Old woman percept. *$p < .05$

Results for group IP

As for the IP group (see Table 3), it was found that semantic congruences were not associated with ocular fixations made in areas of interest (χ^2 (12, N = 142) = 17.519, p = .131). It was noted that gazes at the A3 area were fewer; only 3 fixations were recorded during the audiovisual exposure relating the young woman's voice, where two of the three reports corresponded to the OW percept. On the other hand, there were no fixations in this area during the crossmodal stimulation regarding the old woman's voice.

Ocular fixations in areas of interest 250 ms before percept reports for each body position

As shown in Table 4, there was a difference in the averages of ocular fixations between groups NIP and IP. The result showed that in the position that refers to the vectorial alignment between the G and Zh axes, there was a greater quantity of ocular fixations in the areas of interest analysed with respect to those that occurred in the position corresponding to the vectorial opposition between the G and Zh axes. It was found that the NIP group had more ocular fixations in the areas of interest (M = 4.86; SD = 4.39) compared to the IP group (M = 3.22; SD = 1.89); t (86) = 2.268; p = .0002).

Table 4 Ocular fixations in areas of interest 250 ms before percept reports for each body position

Body position	*n*	*M*	*SD*	*Std. Error Mean*
NIP	44	4.86	4.39	.66281
IP	44	3.22	1.89	.28505

Association between ocular fixations and percepts (NIP group)

The visual percepts that were reported had a statistically significant association with the ocular fixations manifested in the critical bottom-up modulation areas (χ^2 (4, N = 214) = 20.734, p = .0003). Table 5 shows the relation between the reports of the two possible percepts of the bistable image (YW for the young woman percept and OW for the old woman percept) and visual areas of interest (AOIs):

Table 5 Relationship between the reports of percepts and AOIs (NIP group)

AOIs	YW Report	YW Report (%)	OW Report	OW Report
A1	46	79%	12	21%
A2	44	49%	46	51%
A3	5	33%	10	67%
A4	8	40%	12	60%
Background	20	65%	11	35%

Note: Test with Chi-Squared statistic (χ^2 (4, N = 214) = 20.734, p = .0003). AOIs= Areas of interest. YW= Young woman. OW= Old woman.

Table 3 Semantic congruence with each reported percept considering the AOIs (IP group)

AOIs	YW Cong.	% YW Cong.	YW Incong.	% YW Incong.	OW Cong.	% OW Cong.	OW Incong.	% OW Incong.
A1	16	38%	4	15%	8	24%	17	41%
A2	12	29%	14	54%	16	48%	17	41%
A3	1	2%	2	8%	0	0%	0	0%
A4	4	10%	3	12%	1	3%	2	5%
Background	9	21%	3	12%	8	24%	5	12%
Total	42	62%	26	38%	33	45%	41	55%

Note: Test with Chi-Squared statistic (χ^2 (12, N = 142) = 17.519, p = .131). Cong = Congruent; Incong = Incongruent. AOIs= Areas of interest. YW= Young woman percept. OW= Old woman percept.

It was also found that the A1 area significantly favoured the YW percept compared to the other three areas (see Table 6). When reviewing the table of comparisons (Table 5), it is noted that the A1 area, with 79% of the fixations, corresponded to the YW report. As a matter of fact, there was a significant association between this area and the YW percept. Table 6 shows that there was significance ($p < .05$) when comparing the percentage value of fixations in A1 (with report YW) with the other percentage values for each of the other areas of interest that refer to the same YW report. Equally, when comparing ocular fixations in the areas A2, A3 and A4 (regarding OW percept) with fixations made in A1, the results showed that the first three areas were more associated with the OW percept compared with the A1 area.

Table 6 Results of association tests between reports for YW and OW and fixated areas (NIP group)

	AOIs A1 (A)	AOIs A2 (B)	AOIs A3 (C)	AOIs A4 (D)	Background (E)
YW	B* C* D*	-	-	-	-
OW	-	A*	A*	A*	-

Note: Test with statistics from association tests.
*$p < .05$

Association between ocular fixations and reported percepts (IP group)

The percepts did not have a significant association with the ocular fixations in critical bottom-up modulation areas ($\chi 2$ (4, N = 142) = 7.055, p = .133). In other words, there was no evidence to accept an association between ocular fixations and the percepts that were configured during the observation of the stimuli in an inverted position (see Table 7).

Table 7 Relation between reports of percepts and AOIs (IP group).

AOIs	YW Report	YW Report (%)	OW Report	OW Report (%)
A1	33	73%	12	27%
A2	29	49%	30	51%
A3	1	33%	2	67%
A4	6	60%	4	40%
Background	14	56%	11	44%

Note: Test with Chi-squared statistic (χ^2 (4, N = 142) = 7.055, p = .133). AOIs = Areas of interest. YW= Young woman. OW= Old woman.

Discussion

The results showed that, for the NIP group, there was a significant association between the A1 area and the visual report YW when audiovisual integration with semantic correspondence was present. This suggests that the A1 area claims the possibility of perceiving the percept associated with this area, which implies a bottom-up modulation phenomenon. Audiovisual multimodal integration, inferred from the semantic congruences that occurred between audios and visual percepts, is mainly promoted by the recognition of the YW percept while listening to the voice of the young woman, and also while the eyes were fixated on the modulating A1 area. As stated, fixations in area A1 consistent with the YW percept were significantly different when compared to the incongruous percept. On the other hand, the association between the incongruity for the OW percept when looking at area A1 was significantly different when it was compared both with the congruence for the OW percept, and with the incongruence with YW (while viewing area A1). Thus, the semantic congruence that led to audiovisual integration had a connection with ocular fixation areas. The studies conducted by Gale and Findlay (1983) and García-Pérez (1989) refer to regions that can induce the perception of each possible percept of the bistable image used. Indeed, the area that corresponds to the jaw and the mouth of the elderly woman (A3 area) does not favour the YW percept. It implies an encouragement of the OW percept, but not in the way that the young woman's eye area (placed in A1) favours the young woman percept (García-Pérez, 1989). These findings are in line with what was found in the present study, whereby the percentage of reports consistent with the OW (when its modulating area is observed) is not high in relation to the congruences concerning fixations in the modulating area for the young woman report. Moreover, the fact that fixations in areas A2 and A4 did not favour one percept more than the other, is aligned with the findings obtained by Gale and Findlay (1983), in the sense that perception is not influenced by the observation of parts of the image that do not facilitate the recognition of a particular percept (Hsiao et al., 2012; Marroquín-Ciendúa et al., 2020).

As for the NIP group, it was also found that the reported visual percepts had an association with the ocular fixations. It was once again highlighted that area A1 significantly favoured the YW percept in relation to the other three areas. In the case of the area that should modulate the perception of the OW percept, although it descriptively favoured OW (67% of those who had ocular fixations in A3 reported that percept), it did not turn out to be significant. Nevertheless, when comparing fixations in A2, A3 and A4 with fixations in A1 (referring OW percept), a level of significance was set regarding the fact that the first three areas were more associated with the OW percept compared to the A1 area. These results reaffirm the findings provided by Gale and Findlay (1983), particularly in relation to the relevance of area A1 for perceiving the young woman (e.g., García-Pérez, 1989).

As far as the IP group results are concerned, it was found that the congruence between the semantic load of audios and the reported visual percepts were not associated with ocular fixations. It should be considered that the base of reports obtained in relation to eye-fixations in A3 were low (only 3 fixations occurred during the crossmodal stimulation regarding the young woman's voice, where 2 corresponded to the OW percept). There were no significant associations when comparing each area of interest in relation to the

correspondence or non-correspondence with the semantic load provided by each tone of voice. Therefore, it should be considered that there were fewer ocular fixations in the IP group compared to the NIP group. This fact had an effect on the analyses of associations that were carried out. The lower number of ocular fixations manifested in the IP group compared to the NIP could have had a consequence in the ability to make perceptual configurations, because the recognition of the percepts of a bistable image is associated with ocular fixations made in specific areas (Brouwer & van Ee, 2006; Gale & Findlay, 1983; Hsiao et al., 2012). The inversion of idiotropic axes can affect the perception of visual stimuli. It occurs due to the assimilation and accommodation processes that, in turn, entail the incorporation of information alluding to atypical body positions (Balint & Hall, 2015). In addition, compensatory eye movements that emerge as a result of the processing of vestibular information are performed as a reflex act to stabilize the image (Angelaki et al., 2009). In this regard, vestibular-visual interactions imply an intermodal combination of cues while underlying perceptual dimension of heading direction (Yamamoto & Yamamoto, 2006). There is a modulating process through which vestibular signals can exert an influence on visual sensory pathways (Ferrè et al., 2015). This fact encourages corrections or adjustments concerning eye-movements, which, in turn, can affect the capability of making ocular fixations (Angelaki et al., 2009). These issues are involved in the occurrence of unity assumption phenomenon while studying audiovisual perceptual integration based on bistable perception (Marroquín-Ciendúa et al., 2020). Perceptual processes of audiovisual integration might be affected by modulating vestibular information so that ocular fixations come to be involved in the occurrence of the unity assumption.

The findings of the present study have practical implications concerning semantic congruency, an effect that has been stated as a critical factor in multisensory behavioural performance (Laurienti et al., 2004). Unusual body positions have become relevant to the understanding of perception (Balint & Hall, 2015). Therefore, the fact that the direction of idiotropic axes can diminish perceptual performance has been considered as an important issue within the scope of perception for action (Oman, 2003). It implies a constructive nature of perception where perceivers decode stimulation in function of gravitational conditions (Clément & Eckardt, 2005). The relationship between body orientation and perceptual processes has become relevant (Lopez et al., 2008), considering that the perception of reality depends on the integration of the sensory inputs that simultaneously reach different sensory systems (Spence & Squire, 2003).

To close this section, it should be considered that the study outlined here has limitations in terms of its potential. Firstly, the effects of the language (French) were not completely isolated from the effects of the tone, because a few of the vocal sounds may have sounded similar to how they sound in Spanish (the native tongue of all participants). However, they all responded that both audios were incomprehensible. Secondly, some areas of interest were not eye-fixated by IP group as expected. As such, participants of this group looked at the A3 area less in relation to the A1 area, in turn, leading to the dominance of the young woman percept. This fact could partially dismiss the importance of the analysis of association, given the small base of reports in the modulating area concerning the old woman percept. Moreover, as for the statistical power and the effect of size for the Student's t tests (concerning comparisons of eye fixations), the power value was less than .80 *(1-β = .72; d = .48)*. This indicates that there was a probability of 72% of finding statistically significant differences with a medium effect (Cohen, 1992). Therefore, further studies need to be carried out in order to conduct experiments that include more participants, along with more trials which assess ocular fixations as bottom-up modulating factors within the scope of multisensory audiovisual semantic congruency, considering not only several body positions, but also the absence and presence of modulating auditory stimuli. As far as the recognition of percepts is concerned, it is more difficult to identify the two possible percepts of a bistable image in an untypical body position (Clément & Eckardt, 2005). In the present study, when comparing ocular fixations, a significant difference was found in favour of the NIP group. This leads to the fact that perceptual processes of audiovisual integration may be affected by changes in the position of the body in relation to the visual percept that is capable of producing the semantic congruency effect.

Conclusions

Multisensory audiovisual semantic congruency is associated with modulating bottom-up factors, particularly with ocular fixation areas involved in the processing of bistable visual information when idiotropic vectors of human body are aligned with the up direction. Thus, audiovisual multimodal integration, derived from the semantic congruences that are manifested due to the unity that perceptually emerges, can be modulated by ocular fixations made on specific areas of a bistable image. While positioning the body, making the head idiotropic vector point downwards and aligning the trunk idiotropic vector in a 20-degree rotation angle relative to the horizontal, ocular fixations manifested in bottom-up modulating areas are not associated with multisensory audiovisual semantic congruency. Besides, the ability to make eye fixations decreases when the head idiotropic axis is aligned with the gravity vector. Body position can affect visual perceptual processes involved in the occurrence of audiovisual semantic congruency.

Financing

This study was funded by Universidad de Bogotá Jorge Tadeo Lozano.

References

Angelaki, D. E., Klier, E. M., & Snyder, L. H. (2009). A vestibular sensation: Probabilistic approaches to spatial perception. *Neuron, 64*(4), 448-461. https://doi.org/10.1016/j.neuron.2009.11.010

Balint, T. S., & Hall, A. (2015). Humanly space objects — perception and connection with the observer. *Acta Astronautica, 110*, 129-144. https://doi.org/10.1016/j.actaastro.2015.01.010

Bonnet, C. (1994). Psicofísica de los tiempos de reacción: teorías y métodos. *Revista Latinoamericana de Psicología, 26*(3), 431-444.

Brouwer, G. J., & van Ee, R. (2006). Endogenous influences on perceptual bistability depend on exogenous stimulus characteristics. *Visual Research, 46*(20), 3393-3402. https://doi.org/10.1016/j.visres.2006.03.016

Clarke, A. H. (2008). Listing's plane and the otolith-mediated gravity vector. *Progress in Brain Research, 171*, 291-294. https://doi.org/10.1016/S0079-6123(08)00642-0

Clément, G., & Eckardt, J. (2005). Influence of the gravitational vertical on geometric visual illusions. *Acta Astronautica, 56*(9-12), 911-917. https://doi.org/10.1016/j.actaastro.2005.01.017

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155. https://doi.org/10.1037/0033-2909.112.1.155

Ferrè, E. R., Walther, L. E., & Haggard, P. (2015). Multisensory interactions between vestibular, visual and somatosensory signals. *PLoS One*, *10*(4), e0124573. https://doi.org/10.1371/journal.pone.0124573

Gale, A., & Findlay, J. (1983). Eye-movement patterns in viewing ambiguous figures. In *Eye movements and psychological functions: International views* (pp. 145-168). Hillsdale NJ: LEA.

García-Pérez, M. (1989). Visual inhomogeneity and eye movements in multistable perception. *Perception & Psychophysics, 46*(4), 397-400. https://doi.org/10.3758/BF03204995

Gaunet, F., & Berthoz, A. (2000). Mental rotation for spatial environment recognition. *Cognitive Brain Research*, *9*(1), 91-102. https://doi.org/10.1016/S0926-6410(99)00038-5

Haji-Khamneh, B., & Harris, L. R. (2010). How different types of scenes affect the Subjective Visual Vertical (SVV) and the Perceptual Upright (PU). *Vision Research, 50*(17), 1720-1727. https://doi.org/10.1016/j.visres.2010.05.027

Hsiao, J., Chen, Y., Spence, C., & Yeh, S. (2012). Assessing the effects of audiovisual semantic congruency on the perception of a biestable figure. *Consciousness and Cognition, 21*(2), 775-787. https://doi.org/10.1016/j.concog.2012.02.001

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioural performance. *Experimental Brain Research*, *158*(4), 405-414. https://doi.org/10.1007/s00221-004-1913-2

Lopez, C., Lacour, M., El Ahmadi, A., Magnan, J., & Borel, L. (2007). Changes of visual vertical perception: A long-term sign of unilateral and bilateral vestibular loss. *Neuropsychologia*, *45*(9), 2025-2037. https://doi.org/10.1016/j.neuropsychologia.2007.02.004

Lopez, C., Lacour, M., Léonard, J., Magnan, J., & Borel, L. (2008). How body position changes visual vertical perception after unilateral vestibular loss. *Neuropsychologia, 46*(9), 2435-2440. https://doi.org/10.1016/j.neuropsychologia.2008.03.017

Marroquín-Ciendúa, F., Rodríguez, G., & Rodríguez-Celis, H. G. (2020). Modulación de la percepción biestable: un estudio basado en estimulación multimodal y registros de actividad oculomotora. *Tesis Psicológica*, *15*(1), 1-30.

Mittelstaedt, H. (1983). A new solution to the problem of the subjective vertical. *Naturwissenschaften, 70*(6), 272-281. https://doi.org/10.1007/BF00404833

Oman, C. M. (2003). Human visual orientation in weightlessness. In L. Harris & M. Jenkin (Eds.), *Levels of Perception*. New York, NY: Springer. https://doi.org/10.1007/0-387-22673-7_19

Rodríguez-Martínez, G., & Castillo-Parra, H. (2018a). Tareas de búsqueda visual: modelos, bases neurológicas, utilidad y prospectiva. *Universitas Psychologica, 17*(1), 1-12. https://doi.org/10.11144/Javeriana.upsy17-1.tbvm

Rodríguez-Martínez, G., & Castillo-Parra, H. (2018b). Bistable perception: Neural bases and usefulness in psychological research. *International Journal of Psychological Research, 11*(2), 63-76. https://doi.org/10.21500/20112084.3375

Smith, E., Grabowecky, M., & Susuki, S. (2007). Auditory-visual crossmodal integration in perception of face gender. *Current Biology, 17*, 1680-1685. https://doi.org/10.1016/j.cub.2007.08.043

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, *28*(2), 61-70. https://doi.org/10.1250/ast.28.61

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971-995. https://doi.org/10.3758/s13414-010-0073-7

Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology, 13*(13), R519-R521. https://doi.org/10.1016/S0960-9822(03)00445-7

Vatakis, A., & Spence, C. (2008). Evaluating the influence of the "unity assumption" on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica, 127*, 12-23. https://doi.org/10.1016/j.actpsy.2006.12.002

Yamamoto, S., & Yamamoto, M. (2006). Effects of the gravitational vertical on the visual perception of reversible figures. *Neuroscience Research, 55*(2), 218-221. https://doi.org/10.1016/j.neures.2006.02.014

Yeh, S., Hsiao, J., Chen, Y., & Spence, C. (2011). Interplay of multisensory processing, attention, and consciousness as revealed by bistable figures. *i-Perception, 2*(8), 910. https://doi.org/10.1068/ic910