Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009) Jyväskylä, Finland
Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, Päivi-Sisko Eerola (Editors)

Embodied Experience and Communicative Intentions of the Singing Performer

Claudia Mauléon

La Plata's National University-Music Department; and La Plata's State Conservatoire; Argentina

[11]claudiamauleon@cmauleon.com.ar - tecnicavocal@claudiamauleon.com.ar

ABSTRACT

Theories of Embodied Cognition assert that simulation mechanisms underlie inter-subjective communication. On this basis we posit that by solely assessing only visual component of a performance, a naïve audience could make similar judgments to those ones elicited by audiovisual or aural perception. Five vocal performances by performers of different levels of expertise were assessed using various of perception (audiovisual, visual and aural perception) by 90 musically uneducated subjects randomly assigned to a specific modality. Subject's task consisted of pronounce an aesthetical judgment of the performances using an 11-point scale. Results assessed by ANOVA test of repeated measures showed significant differences between the factors SINGERS (F 16,296 p< 0,000) and CONDITION (F 8,622 p< 0,001) meaning that singers were judged differently among them and that, judgments were quantitatively different through each perceptual modality. Instead, factors' interaction (SINGERS x CONDITION) was non significant (F 1,090 p<0,372) indicating that each singer was similarly evaluated via the three perceptual modalities. Consequently, results support the idea of a cross-modal correspondence in the reception of vocal performance. Besides, the lack of knowledge of the lyrics and the style of the musical piece on the part of the audience, allows us to suppose that judgments were based on sensory-motor simulations.

I. BACKGROUND

A. The Communicative Intentions of the Interpreter

Theories of embodied cognition posit that mental representations are the images with which our mind infers, builds concepts and establishes interactive relationships with the environment. Thus mind and body are a whole and the perceptions and actions of the individuals are entwined with their cognitive construction of the world (Di Paolo, E.A., 2005; Maturana, H.R., 1981; Metzinger, T., 2004; Niedenthal, P.; Barsalou, L.; Winkielman, P. et al., 2005). From this perspective, we see the body-mind of the singer as an integrated whole that shapes in vocal sounds the musical representations inside the mind of the singer.

During the education of the singer two situations occur in parallel with one another. Firstly, a new scheme between voice and body is gradually established. This scheme is increasingly well differentiated from those ones specific to speech or swallowing. Secondly, musical skills are developed. As a result, musical abilities and vocal-body scheme for singing become progressively intertwined during educational period of the singer (Deutsch, D., 1999; Mauléon, C., 2004; 2007; Mauléon, C.y Gurlekian, J., 2001; Mauléon, C.; Pessolano, F.; Gurlekian et al., 1999; Miklaszewski, K., 2004; 1992; Sloboda, J.A., 1985; 1997). This means that the singer's musical representations and multi-level perceptual-motor coordinations gradually become harmonically integrated. However, the extent and fluency of such integration would depend on the accuracy of the singer's mental representations of musical structure and sound, and/or the degree of development of the singer's vocal-performance scheme. Both skills could not be present at the level; therefore expertise, could be seen as an elevated integration between musical and sensory-motor schemes, in which the body-mind of the singer flows with the musical stream (Sloboda, J.A., 1997).

Briefly, as Edlund (1997) pointed out, the mental image the musicians construct of a musical piece is strongly mediated by the configuration of motor and proprioceptive patterns. These patterns are governed by the relationship between the limits of the human anatomy and the constructive and acoustic features of the instrument. Thus, the anatomic and functional relations within the vocal instrument are the potential and physical limit of performer's accomplishment in singing. Besides, as in every instrumental performance, singing implies a series of movements that lead to complex actions that allow the materialization of the imagined sound. Thus musical thoughts, performance gestures, and the idiosyncratic gestuality of the interpreter are knitted together as complex motor patterns

In the light of the previous ideas, we see the voice in its artistic use as the result of a dynamic interplay between the *corporal reality* and *the communicative intentions* of the interpreter. From this view we understand the term *corporal reality*, as a concept encompassing both the anatomic and functional aspects beneath the quality of the vocal sound and the corporal situation of the interpreter in his/her relationship with the environment, both mediate and immediate. The *communicative intentions* of the interpreter are thus the conscious and unconscious musical ideas conducting the performance. The *communicative intentions* include affective or emotional features, which give rise of a complex imagery wherein body situation (*corporal reality*) plays a role.

B. Cross-modality in Singing Perception

The most natural context in which to experience a performance is through audiovisual perception in real time, where the audience can listen to the musical sounds and see the gestures which originate them. Nevertheless, since the advent of sound's recording and reproduction technologies, the auditive modality is the most frequently used means of accessing a performance. In both cases, people enjoy the music and they can distinguish between different interpretations of the same piece. It seems that, suppressing the vision of the gestures at the origin of the sound does not prevent the audience from perceiving the *communicative intentions* of the interpreter. Apparently the *communicative intentions* of a performance are sufficiently present in the sound itself.

These ideas are supported by theories which propose the existence of cross-modal properties in perception(Meltzoff, A.N., 2007; Meltzoff, A.N.y Borton, R., 1979; Meltzoff, A.N.y Brooks, R., 2007; Stern, D., 1998; , 2000; , 2004; Stern, D.; Hofer, L.; Haft, W. et al., 1998; Trevarthen, C., 2000; Trevarthen, C., 2004), and by theories related to the existence of simulation mechanisms linking perception, communication and empathy (Gallagher, S.y Meltzoff, A., 1996; Gallese, V., 2000; Metzinger, T., 2004).

As in every performance, singing is guided by sensory-motor perceptions. Tactile proprioceptions (deriving from the vocal tract, and oral-facial innervations), kinetic ones (coming from posture, general body movements, movements of articulators, etc.) and the perceptions coming from the ear (both internal and external) all of which provide feedback to the singer.

In turn, the spectators have access to the sound and the external movements of the performance. According to the theories relative to embodied cognition, mechanisms such as mirroring and cross-modal sensory mapping (Maturana, H.R., 1981; Metzinger, T., 2003a; , 2004b; , 2007; Metzinger, T.y Gallese, V., 2003; Niedenthal, P.; Barsalou, L.; Winkielman, P. et al., 2005) allow a witness to a performance to project clues contained in sound and movements into his/her own body. Thus, the spectator would simulate in his/her own body-mind, sensory-motor experiences that would trigger images about the *communicative intentions* of the performer. Moreover, these images would be influenced by the personal and cultural backgrounds of the audience.

In this context we could ask ourselves if the *communicative intentions* of the musical interpreter are also present in the dynamics of gestures “per se”. That i.e. could the audience infer the *communicative intentions* of the musical interpreter solely by means of gestures?

Antecedents in different fields of musical studies tend to support such an idea. For instance Davidson (1993) showed that an audience could through the visual output distinguish between deadpan, exaggerated and regular performances. Besides, other studies produced on the subject have pointed out that gestures chosen by expert performers are mainly conditioned by expressive aspects and not by technical considerations alone (Cadoz, C.y Wanderley, M.M., 2000; Friberg, A., 2004; Godøy, R.I., 2004; Godøy, R.I.; Haga, E. yJensenius, A.R., 2006; Miklaszewski, K., 2004; Parncutt, R., 1997; Wanderley, M.M., 1999). This last group of papers concur with the ideas proposed by Imberty (1997) and Sloboda (1997) remarking that experts organize their interpretation of a piece and the sound matter in a very different way than the way in which novices do. In other words, the actions of the interpreter correspond with his/her internal configuration of the musical structure and the imagined sound.

Resultantly one could infer that gestures “per se” would transmit the very *communicative intentions* of the performer, and thus gestural quality would vary significantly among experts and novices.

II. HYPOTHESIS AND AIMS

C. Rationale

In the accordance with the basis of the background put forward, the following reasoning ensues:

- In a singing performance, the vocal sounds and the corporal gestures are the resulttor actions led by the *communicative intentions* of the singer.
- Singers’ mastery level and music configuration play a role in the way performers organize and communicate their interpretation.
- Therefore, it could be stated that the general features of an interpretation can be grasped from the different perceptive moods involved in its perception.

Based on these ideas, was designed and experiment for the purpose of testing the incidence of each perceptual modality on the aesthetical judgements of a naïve audience.

D. Objective

The objective of the experiment was to observe the incidence of audiovisual, visual and aural modalities of perception on the judgements of a naïve [1] audience with regards to five singing interpretations of different level of expertise.

E. Hypothesis

H1. Gesture and sound separately, and gesture and sound combined, will account for the level of expertise of the performances.

H0 Gesture and sound separately, and gesture and sound combined, will not account for the level of expertise of the performances.

F. Prediction

We predicted the following tendencies in data:

(a) Punctuation of subject’s aesthetic judgments will correspond to the variations in the level of expertise of the performances; thus, the more expert the performance, the higher its score.

(b) Audiovisual, visual and aural modalities will be considered in different ways by the audience. Thus the different modalities of perception will qualify quantitatively as different.

(c) The patterns of the judgments relative to each performance will be sustained throughout the three modalities of perception. Thus the performance which qualifies higher or lower in one modality will follow suit in the other two.

III. METHOD

We adopted an experimental model proposed by Davidson (1993) and Vines et al. (2004) —in which performances were judged from three different sensorial experiences, namely audiovisual, aural and visual.

Subjects: 90 students males and females, of an average age 22, with no systematic music education and no knowledge or familiarity with the style, or with the piece involved in the experiment, or the language of the lyrics. The subjects were

[1] We define a “naive audience” as one composed by individuals with no systematic knowledge of music or familiarity with the musical piece or style at stake.

randomly assigned to three different groups of 30 individuals each (experimental conditions 1, 2 and 3).

Performances: Five singers collaborated with the study, each recording their performances of an opera aria. The recordings were made while singers sang over the recorded orchestral accompaniment. The singers had different levels of skilfulness. Two of them, S1 and S2 were students in their second and third year of training respectively; S3, S4 y S5 had finished their studies. S3 had another profession and only sang occasionally. In contrast, S4 y S5 worked at a professional opera chorus; furthermore S5 was at the initial steps of a solo carrier.

A panel of professional musicians (n=5) ranked the five interpretations. The level of agreement amongst the experts in percentage terms was 68%. The agreement for any of the five performances were: 100% of agreement for the best interpretation, 80% of agreement for the second, and 60 % of agreement for the third and the poorest interpretations, the less percentage of agreement (40%) was found for the fourth position in the ranking. Performances resulted ordered as follows from the better to the poorest one S5, S4, S3, S2, and S1.

Stimuli: The musical piece selected was the aria *O mio Babbino Caro* by *G. Puccini* from the opera Gianni Schichi. The stimuli were constructed from five video recordings especially taken for the purpose of this test (see performances). A fragment of 1,57 min. (bars 4 to 31) was processed in order to compose three kinds of stimuli: (a) audiovisual output of the five interpretations; (b) aural only output of the five interpretations; and (c) visual only output of the five performances.

Experimental design: The experiment included three experimental conditions. For condition (1) participants assessed the interpretations through audiovisual perception; for condition (2) the audience experienced the performance only by aural perception only; and participants in condition (3) assessed the performances through the visual output only.

Procedure: Subject's task consisted of producing an aesthetic judgment for each performance through a scale of 0 to 10 points (meaning 0 poor and 10 excellent). Participants assessed the performances in a random order separated each by 5 seconds on which they should mark the estimated punctuation in a grid. The test lasts approximately 10 minutes.

Equipment: The test was administrated through a PC notebook Acer aspire 5670 and a projection device View Sonic 2000 lumes (conditions 1 and 3); amplifier Ibañez 15 wats, 4 Ω (conditions 1 and 2) and projection screen View Sonic 80 inches (conditions 1 and 3). Video recordings were processed with Adobe Premier Pro 2.0 software.

IV. RESULTS

Results were assessed by ANOVA test of repeated measures, with the factor "SINGER" as intra-subjects' factor and "CONDITION" as between subjects' factor. Both factors showed significant differences (F= 16, 296 p<0,000 y F=8,622 p<0,001 respectively). The interaction between both factors ("SINGER X CONDITION") was no statistically significant (F=1,090 y p>0,372) (Table 1).

Table 2 summarizes details of the mean scores, and their deviation obtained for each interpretation from each experimental condition. Note that mean scores are lower and deviations higher in condition (3). Nevertheless, the general tendency is sustained through the three experimental conditions (Gaphic 1).

Graphic 1 also shows that audiovisual perception (condition 1) scores higher than conditions 2 (visual perception) and 3 (aural perception).

Graphic 2 represents the way the panel of experts ranked the five interpretations, if this information is compared it with the answers given by the audience, S3 shows an interesting difference.

The significant difference in factor 'SINGERS' indicates that the audience qualified differently the five performances (graphic 1) and that these differences are sustained throughout the three experimental conditions.

Table 1 shows that the five fingers were judged differently; that there were significant differences in scores between experimental conditions; and that the interaction between the factors, 'singer x experimental condition', was not significant.

	Sum of squares	gl	Quad. mean	F	Sig.
Singer	102,01	4	25,50	16,29	,000
Condition	98,96	2	49,48	8,622	,001
Singer x Condition	13, 58	8	1,69	1,090	,372

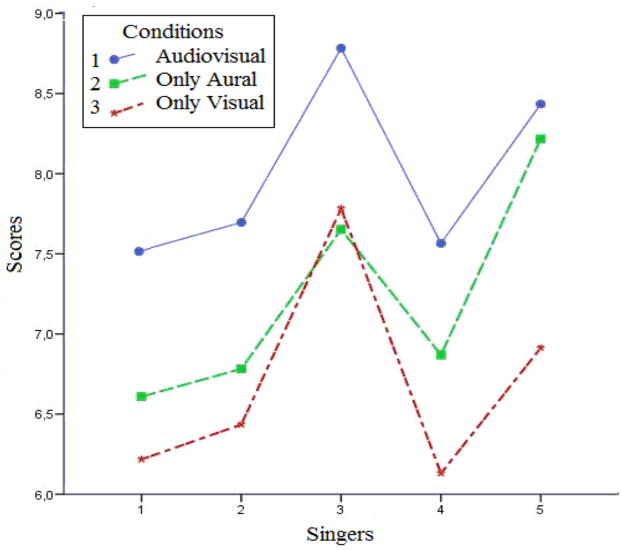
Graphics 1 and 2 allow for the comparison between the judgments of the audience and those of the expert panel. With the exception of S3, the judgments of the audience were similar to those of the experts. These results allow prediction (a) to be accepted as valid.

Table 2: Shows the mean scores and their deviations for five performances assessed through 3 different conditions of perception (1) audiovisual; (2) aural, and (3) visual.

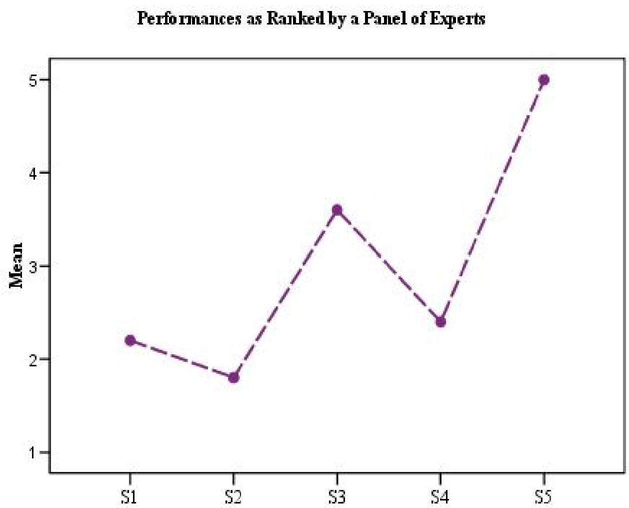
Descriptive Statistics			
Experimental Conditions	Singer	Mean	Deviation
1 Audiovisual	1	7,52	1,81
	2	7,70	1,34
	3	8,78	0,95
	4	7,57	0,99
	5	8,43	1,34
2 Aural	1	6,61	1,34
	2	6,78	1,08
	3	7,65	1,23
	4	6,87	1,25
	5	8,22	1,16
3 Visual	1	6,22	1,88
	2	6,43	1,99
	3	7,78	1,78
	4	6,13	2,58
	5	6,91	2,13

The significant difference in factor 'CONDITION' indicates that visual, audiovisual and aural modalities were punctuated quantitatively different and therefore, it confirms prediction (b) as valid. These are very logical results if one considers that the most ecological way of appreciating a musical performance is audiovisual modality, and that the visual one is a completely artificial way of asses a musical phenomenon.

On the other hand, the non significant difference in the interaction between factors ´SINGER x CONDITION´ indicates that the way, performances were judged was consistent throughout the three different modalities of perception. In other words, the perceptual modality of perception does not modify the way in which a particular performance was judged. As these judgments were coherent with the level of expertise of the singers, we can say that the modality of perception of the performances does not influence the judgment related the level of expertise of each singer. These results are in agreement with prediction (c).



Graphic 1: shows the mean scores obtained by five performances assessed through three different experimental conditions (1) audiovisual perception, (2) aural only perception and (3) visual only perception. Note that the general tendency of the punctuations is sustained throughout conditions.



Graphic 2 shows the ranking of the five performances estimated by a panel of expert musicians. Note the discrepancy relating S3 and S5 with the judgments produced by the participants of the experiment.

On one hand, as is shown on table 2 and graphic 1, mean scores for audiovisual modality (condition 1) were higher, indicating that gesture-sound interaction influences positively on the judgments, since *singers' communicative intentions* were conveyed in a multimodal way. Thus the voice and the gestures would compose a complex that would make clearer musical and expressive information for the audience. In addition, Graphic 1 seems to show the existence of a strong impact of gestures on the punctuation of performance S3 (Note that compared with the group, this singer showed the higher mean for condition 1). Considering that S5 was evaluated as the best performance by the experts, we can posit two questions: (a) were these differences significant? If so, (a) are they determined by: the sum of visual and aural channels of information, or (b) are they the consequence of a multimodal complex gesture-sound? Analysing the data of performances S5 (the best) and S3 (the third in the group according to the experts) we found that, while differences were significant for condition 1 (audiovisual) and 2 (aural), they were not for condition 3 (visual) (See table3). In other words, the significant differences obtained by the performances in condition 1 (audiovisual) would not be explained by the sum of visual and aural channels.

Table 3 shows the ANOVA test comparing the scores of S3 and S5. It indicated that differences between singers were significant for conditions 1 and 2, but not for condition 3/ the significant differences obtained by the performances in condition 1 (audiovisual) would not be explained by the sum of visual and aural channels.

ANOVA		Sum of Squares	df	Mean Square	F	Sig.
Condition 1 Audiovisual S3 vs. S5	Betw.Groups With.Groups Total	11,971 7,942 19,913	5 17 22	2,394 ,467	5,125	,005
Condition 2 (Aural) S3 vs. S5	Betw.Groups With.Groups Total	34,325 24,361 58,686	6 28 34	5,721 ,870	6,575	,000
Condition 3 (Visual) S3 vs. S5	Betw.Groups With.Groups Total	13,871 76,639 93,500	7 24 31	1,980 3,318	,597	,752

Consequently, we cannot say, *prima facie*, that higher scores in condition 1 (audiovisual) would be due to the sum of gesture and sound components. This first analysis seem to support the idea of a of a gesture-sound complex..

We estimate that our results allow to reject H0 and accept H1 as a plausible alternative; this is, that gesture and voice separately, and gesture and voice together account for the level of expertise of the performances.

V. CONCLUSIONS

The objective of the experiment was to observe the incidence of visual, aural and audiovisual modalities of perception on the judgments of a naive audience about five performances each of different level of expertise. The results have shown that the audiovisual modality has a positive incidence on the judgments: that means the scores in this condition were higher than those of the other two (aural and visual). This could apply to two interpretations of the phenomenon: a hypothesis of addition (audiovisual perception is the sum of aural and visual perception), or a hypothesis of multimodal complexity (gesture and voice are a complex unique and indivisible whole, that allows the spectator a deeper experience of the musical event). The addition hypothesis has been supported by some studies (Vines, B.; Wanderley, M.M.; Nuzzo, R. et al., 2004), nevertheless in our work it is not sustained by the statistical analysis (table 3). From these performances an important difference in the gesture component that seemed to be reflected in the audiovisual one is observed. Nevertheless, the statistic test showed that the observed differences were not significant. Instead, the curve for aural perception resulted significantly different; this fact would indicate that when only listening to S5, this performance was considered better than S3. Therefore it could be inferred that the increasing in the scores for audiovisual reception in performance S3, is due to the interaction of gestures and sounds and not to the addition of both components.

Finally, we should consider the fact that the participants in our experiment had no formal music training and were not familiar with the piece, the style, or the lyrics, and they couldn’t grasp the meaning of the poem as it was sung in a foreign language. Consequently we can state that the audience aesthetic judgments were based exclusively in their sensory perception and so it is feasible that their appreciation of the performances would be led by sensory-motor simulations,as is proposed by some scholars.

We estimate that a detailed study of the components of gesture and sound in conjunction with the musical structure, would allow the development of new hypotheses relating to the way gestures and sounds are intertwined conforming a whole indivisible complex.

VI. REFERENCES

Cadoz, C. & Wanderley, M. M. (Eds.). (2000). Gesture - Music. Paris: Ircam - Centre Pompidou.

Davidson, J. W. (1993). Visual perception of Performance manner in the Movements of Solo Musicians. Psycholgy of Music, 21, 103-113.

Deutsch, D. (1999). The Psychology of Music. San Diego: Academic Press.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. Phenomenology and the Cognitive Sciences, 4, 429-452.

Friberg, A. (2004, 28-30). In J. S. B. Brunson (Ed.), A fuzzy analyzer of emotional expression in music performance and body motion. Paper presented at the Proceedings of Music and Music Science, Stockholm.

Gallese, V. (2000). The Inner Sense of Action. Agency and Motor Representations, Journal of Consciousness Studies (Vol. 7, pp. 23–40).

Gallese, V. (2003). The manifold nature of interpersonal relations: The quest for a common mechanism. Philosophical Transactions of the Royal Society, B (358), 517–528.

Gallese, V. (2007). The “Conscious” Dorsal Stream: Embodied Simulation and its Role in Space and Action Conscious Awareness, PSYCHE (Vol. 13, pp. 1-20).

Gibbs, R. W. (2006). Embodiment and Cognitive Sciences. Cambridge: Cambridge University Press.

Godøy, R. I. (2004). Gestural Imagery in the Service of Musical Imagery. In A. Camurri y G. V. GW (Eds.), LNAI (Vol. 2915, pp. pp. 55–62). Berlin Heidelberg Springer-Verlag

Godøy, R. I.; Haga, E., y Jensenius, A. R. (2006). Playing “Air Instruments”: Mimicry of Sound-producing Gestures by Novices and Experts. In Springer (Ed.), Libro Gesture in Human-Computer Interaction and Simulation (Vol. Volumen Volume 3881, pp. 256-267). Berlin / Heidelberg

Imberty, M. (1997). Qu'est-ce que le "Mouvement" d'une Ouvre Musicale? In F. Escal y M. Imberty (Eds.), La Musique au regard des sciences humaines et des sciences sociales (Vol. I). Paris: L'Harmatan.

Jeannerod, M. (1994). The representing brain: New correlates of motor intention and imagery. Behavioral and Brain Sciences, 17, 187-245.

Maturana, H. R. (1981). Autopoiesis. In M. Zeleny (Ed.), Autopoiesis: A theory of living organization. . New York: North Holland Publishers.

Mauléon, C. (2004). In O. Musumeci (Ed.), Learning to be a Singer Teacher: A new profie of singing voice tuition (pp. 137-148). Paper presented at the Isme 's Ceprom Biannual Meeting. Preparing Musicians Making New Sounds Worlds, Barcelona. Escola Superior de Música de Catalunya.

Mauléon, C. (2007). ¿Es la voz un instrumento musical? .

Mauléon, C., & Gurlekian, J. (2001). In F. Shifres (Ed.), Consonantes oclusivas sordas en el canto. Un estudio sobre la /t/ (pp. Formato CD). Paper presented at the Primera Reunión de Saccom. La Música en la Mente. Procesos implicados en la experiencia musical., Avellaneda, Argentina. Saccom.

Mauléon, C.; Pessolano, F.; Gurlekian, & De Vito, E. (1999). Acoustic, Perceptual, Air volume and Transdiaphragmatic Pressure Measurements in a Supported Singing Voice, 28th Annual Symposium Care of the Professional Voice. Philadelphia: Singular publisher Group.

Meltzoff, A. N. (1979). Intermodal matching by human neonates Nature, 282(5737), 403-404.

Meltzoff, A. N. (1999). Born to Learn: What Infants Learn from Watching Us. In N. Fox y J. G. Worhol (Eds.), The Role of Early Experience in Infant Development (pp. 1-10): Skillman, NJ: Pediatric Institute Publications, 1999.

Meltzoff, A. N. (2005). Imitation and Other Minds: The "Like Me" Hypothesis. In S. H. a. N. Chater (Ed.), Perspectives on Imitation: From Neuroscience to Social Science (Vol. Vol. 2, pp. 55-77). Cambridge, MA: MIT Press.

Melzoff, A., y Brooks, R. (2001). "Like me" as building block for unerstanding other minds: Bodily acts attention and intention. In B. Malle; L. Moses y B. D. (Eds.), Intentionsa and intentionality: Foundations of social cognition. Cambridge, M.A.: Mit Press.

Metzinger, T. (2004). Being no One. Cambridge, Massachussetts: Mitt Press.

Miklaszewski, K. (2004). What and why do we need to know about music psychology research to improve music instrument teaching? In W. D. Jane (Ed.), The Music Practitioner. Research for the Music Perormer, Teacher and Listener. Lincolnshire, UK.: Ashgate.

Niedenthal, P.; Barsalou, L.; Winkielman, P.; Krauth-Gruber, S., y Ric, F. (2005). Embodiment in Attitudes, Social Perception and Emotion. Personality and Social Psychology Review, 9(3), 184-211.

Parncutt, R. (1997). In A. Gabrielsson (Ed.), Interdependence of right and lef hands in sight-read, written, and rehearsed fingerings of piano music paralel octaves (pp. 702-705). Paper presented at the 3rd. Triennial Escom Conference, Upsala, Sweden. ESCOM.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. . Trends In Neuro Science, 21(5), 188- 194.

Sataloff, R. (1992). The Human Voice. Scientific American, 267(6), 108-115.

Sloboda, J. A. (1985). L'Esprit Musicien. La psychologie cognitive de la musique. Bruxelles.

Sloboda, J. A. (1997). Pericia Musical. Orpheotron, 1, 7-9.

Vines, B.; Wanderley, M. M.; Nuzzo, R.; Levitin, D., & Krumhansl, C. (2004). Performance Gestures of Musicians: What Structural and Emotional Information Do They Convey? In Springer (Ed.), Gesture-Based Communication in Human-Computer Interaction (Vol. 2915). Berlin / Heidelberg.

Wanderley, M. M. (1999). Non-obvious Performer Gestures in Instrumental Music. In Lecture Notes In Computer Science: Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction (Vol. 1739, pp. 37 - 48). London, UK Springer-Verlag.