UNIVERSITY OF ROME ‘LA SAPIENZA’

Department of Psychology

European PhD course in “Cognitive plasticity and rehabilitation”

cycle XXV

FINAL DISSERTATION

“The sound of actions: a Mismatch Negativity (MMN) study”

Candidate: Luigi Grisoni

Supervisor

Professor Salvatore Maria Aglioti
Social and Cognitive neuroscience Laboratory
Department of Psychology
Sapienza -University of Rome

Reviewers

Prof. Friedemann Pulvermüller
Institute of German and Dutch Studies
Department of Philosophy and Humanities
Freie Universität, Berlin

Prof. Alessio Avenanti
Centre for studies and researches
in Cognitive Neuroscience
Faculty of Psychology
University of Bologna ‘Alma Mater Studiorum’

Prof. Patrizia Pantano
Department of Neurology e Psychiatry
Sapienza -University of Rome

Official Referee

Prof. Friedemann Pulvermüller

TABLE OF CONTENTS

Introduction

Perception and representation of actions are a multimodal experience modulated by visual (Haslinger *et al.*, 2005), auditory (Kaplan *et al.*, 2007), somatic (Avenanti *et al.*, 2007), and even olfactory inputs (Rossi *et al.*, 2008). The discovery in the monkey brain of double-duty neurons involved in action execution and action observation (visuo-motor mirror neurons) provided a new insight into the possible parity between action perception and action execution (Buccino *et al.*, 2001; Decety *et al.*, 1997; Fogassi *et al.*, 2005; Gallese *et al.*, 1996; Gazzola and Keysers 2009; Rizzolatti *et al.*, 1996). Studies indicate, for example, that mere observation of an action may strengthen the motor representation of the observed action (Stefan *et al.*, 2005) and that mere motor experience of a particular action may improve its visual discrimination (Casile and Giese 2006). In view of these findings and possibly also of the notion that vision dominates the other senses and leads virtually any type of cross-modal and perceptuo-motor interaction, several studies have focused on the link between action observation and action execution (Buccino *et al.*, 2004; Pazzaglia *et al.*, 2008; Rizzolatti *et al.*, 1996; Vogt *et al.*, 2007). However, non-visual senses are fundamental for interacting with the environment when vision is not available. In addition to conditions when sight is absent (Amedi *et al.*, 2005; Pietrini *et al.*, 2004; Ricciardi *et al.*, 2009), several daily life behaviors are guided by non-visual senses. Each action of animals and humans produces a characteristic sound that may permit its unequivocal recognition.

In the first part of this Introduction I will discuss animal studies, which provide direct evidence that action is inherently linked to multisensory cues, as well as the studies carried out on humans demonstrating the link between “resonant” fronto-parietal-temporal networks and the ability to represent an action by hearing its sound. In particular, I will examine these evidences in favour of somatotopy as a possible representational rule underlying the auditory mapping of actions. In the second part I will build the hypothesis tested in the following chapters. I will connect the *somatotopy rule*, above mentioned, within the framework of Auditory Scene Analysis (ASA). This link will be develop throughout the thesis and will be discussed on different levels (i) the functional role of a such ASA-like mechanisms in the action perception (ii) neurophysiological evidences for connecting the two research fields (i.e. somatotopy vs tonotopy) (iii) experimental evidences indicating analogous mechanisms underlying the two processes.

Chapter I

Mechanisms and neural underpinnings of audio-motor action mapping

1.1 Direct evidence in animals

The classical view that posterior and anterior brain areas are involved in sensory and motor processes, respectively, has been challenged by pioneering single-cell recording studies on the brain of macaque monkeys who observed the actions performed by monkey or human agents ("mirror neurons", see Rizzolatti and Craighero 2004 for a review). These studies reveal that viewing the actions performed by another agent may trigger the activity of the frontal and parietal cortical neurons, which are also involved in both the execution and planning of goal-directed actions (di Pellegrino *et al.*, 1992; Fogassi *et al.*, 2005; Gallese *et al.*, 1996). The existence of double-duty "mirror" neurons in classical motor areas indicated, for the first time, that a common neural substrate may underlie the perceptual and motor aspects of actions. Remarkably, some of these double-duty, frontal mirror neurons in fact turned out to be triple-duty cells, activated not only by action observation and execution but also by the acoustic perception of action-related sounds (Kohler *et al.*, 2002). Moreover, these cells can display a multimodal additive response, suggesting they are involved in complex audio-visuo-motor integration functions (Keysers *et al.*, 2003). Note, however, that some of these complex perceptuo-motor neurons respond strongly to action sounds (e.g. paper tearing), but not to the observation of actions (e.g. grasping of an object) performed by another human or monkey agents thus indicating their audio-motor specificity (Keysers *et al.*, 2003).

1.2 Transcranial magnetic stimulation studies

Although direct evidence of the existence of mirror neurons in humans is still lacking, increasing numbers of studies have provided indirect proof of motor mirroring in humans. In a seminal study on this issue, single-pulse transcranial magnetic stimulation (TMS) was delivered to the primary motor cortex of healthy subjects while they observed a model grasping an object. The study revealed that mere observation of the action induced an increase in motor evoked potentials (MEPs) recorded from the same muscles that would be active during actual execution of the same movements (Fadiga *et al.*, 1995). This increase in MEP modulation during action observation was expanded by subsequent single-pulse TMS studies (Candidi *et al.*, 2010; Gangitano *et al.*, 2001; Romani *et al.*, 2005; Strafella and Paus 2000, Urgesi *et al.*, 2006, 2010). In addition, listening to action-related sounds seems to increase cortico-spinal reactivity. Indeed, MEPs recorded from hand muscles during passive listening to sounds associated with bimanual actions produced greater cortico-spinal facilitation than control sounds or sounds associated with leg movements (Aziz-Zadeh *et al.*, 2004).

1.3 Functional magnetic resonance imaging studies

Clear evidence for selective involvement of brain regions in sound-into-action translation processes has been provided in the past few years by the utilization of functional magnetic resonance imaging (fMRI) studies. The first evidence for the existence of an audio-motor mirror system in humans was obtained by asking healthy subjects to passively listen to hand- or mouth-related sounds (Gazzola *et al.*, 2006). The dorsal premotor cortex and inferior parietal lobe in the left hemisphere showed a stronger response to the sound of a hand action than to the sound of a mouth action. In contrast, in the left ventral premotor cortex, the response to the sound of a mouth action was stronger than to the sound of a limb action. Interestingly, the audio-motor mirror system largely responded to the sight of similar actions, thus hinting at the multimodal nature of action simulation. Galati and coworkers (Galati *et al.*, 2008) performed an event-related fMRI study in which a hand or mouth action-related sound (or a sound related to an environmental event) was or was not congruent with a previously presented written word that acted as an unconscious cross-modal priming stimulus. A clear neural signature of prime congruency specific for action sound trials was found in the left inferior frontal and posterior temporal regions, supporting the notion of a fronto-parietal network underlying audio-motor mirroring. Interestingly, the fronto-parietal network is activated during the execution, viewing, and hearing of hand movements, supporting the inherently multimodal nature of motor mirroring (Galati et al 2008, Gazzola et al 2006).

It is noteworthy that environmental sounds not associated with human actions (Galati *et al.* 2008), the phase-scrambled versions of action sounds (Gazzola *et al.*, 2006), and silent events (Lewis *et al.*, 2006), did not activate any of these areas. Because acting upon environmental elements (e.g. grasping rain, wind, or fire) is not possible, the motor system may not be involved in the perception of the sounds related to these elements. Therefore, environmental sounds do not generally qualify as sounds that are associated with embodiment.

1.4 Evoked potentials and Magnetoencephalography studies

The time course of action-related sound processing was explored using multi-channel event-related potentials (ERPs) in an elegant study using a visuo-auditory version of the repetition suppression paradigm, where written words could or could not be semantically related to sounds evoked by human hand (e.g. clapping), or mouth (e.g., whistling) actions, or by other non-human activities (e.g., the sound of the rain; Pizzamiglio *et al.*, 2005). In this study, the left posterior superior temporal and premotor areas were selectively modulated (peak at 280 ms) by action-related sounds, while the left and right temporal poles were modulated (peak at 300 ms) by non-action-related sounds. This pattern of results clearly indicates that separate neural system are used to represent sounds, which can or cannot be translated into human actions. Electrical neuroimaging analyses of auditory evoked potentials in response to sounds that typically cue a responsive action (e.g. a door bell), versus sounds that do not elicit automatic motor responses (e.g. a church bell), revealed the modulation of neural activity in the premotor and inferior (pre)frontal regions, mainly on the left side (De Lucia *et al.*, 2009). In keeping with the study by Pizzamiglio *et al.* (2005), the effects peaked at about 300 ms, well after the general categorical object discrimination, thus supporting the notion that the reported electric modulation is associated with audio-motor action mapping. Of particular interest, for the works that will be presented in the next chapters, is the investigations of the sound-into-action translation process through the mismatch negativity (MMN). At this regards, Hauk and coworkers presented to participants a multideviant paradigm in which the deviant stimuli were finger and tongue clicks as well as sounds with comparable acoustic features but no association with actions (Hauk *et al.*, 2006). Action-related sounds induced larger mismatch negativity than control at ~100 ms after the stimulus presentation. Topographical distribution analysis showed that hearing sounds related to finger and tongue actions induced higher neural activity in left hemisphere motor areas as well as in the more inferior regions of both hemisphere, respectively (Hauk *et al.*, 2006). The application of the same experimental stimuli, and a paradigm of subdural EEG recording in a 12-year-old girl undergoing intracranial monitoring of epileptic seizures, provided an accurate picture of the time course of the sound-into-action translation process (Lepage *et al.* 2010). Neural activity recorded from electrodes overlying the functionally defined hand representation of the motor cortex revealed both early (100 ms) and late (250 – 450 ms) modulation effects of natural finger-clicks when compared with control sounds. Although coming from a single subject, the results hint at the presence of two distinct time windows of M1 activation after action sound are heard and suggest that both early (Hauk et al, 2006) and late (De Lucia *et al.*, 2009; Pizzamiglio *et al.*, 2005) resonance may be triggered by hearing action sounds. One particular Magnetoencephalography (MEG) study has specifically tested the effect of hearing action-related sounds on motor cortex activity, providing evidence of audio-motor mirroring in humans (Caetano *et al.*, 2007). In particular, the authors of this study explored the modulation of the 20-Hz rhythm. The 20-Hz activity is typically suppressed during movement performance and reappears soon after. This phenomenon is known as rebound-effect and reflects the stabilization of the primary motor cortex after a perturbation. A rebound of 20-Hz activity was found not only after execution and visual observation of actions, but also after hearing the sound associated with the same action, demonstrating the clear influence of vision and audition on action mapping (Caetano *et al.*, 2007).

1.5 Representational rules underlying the auditory mapping of actions

As previously discussed, different cognitive neuroscience techniques and experimental protocols in healthy subjects have provided convergent evidence for the existence of a fronto-parietal network involved in audio-motor mapping of human actions. Audio-motor mirroring is modulated by several factors including the intention to act (Knoblich and Repp 2009) the presence of body-object interactions (Lewis *et al.*, 2005; De Lucia *et al.*, 2009), and the body part involved in the action evoked by the heard sound (Fadiga *et al.*, 2002; Pizzamiglio *et al.*, 2005; Hauk *et al.*, 2006; Gazzola *et al.*, 2006; Pazzaglia *et al.*, 2008; Galati *et al.*, 2008). In particular, the latter factor involves the specificity of the bodily instantiation of cognitive operations, which is known as "embodiment" (Barsalou, 2008). It is widely believed that performance of actions implies the activation of body parts according to a somatotopic, homuncular cortical representation (Penfield and Rasmussen, 1950) (Figure 1.1). Studies of the visual observation of actions suggest that somatotopy may also be an important representational rule for mirroring processes. Indeed, neural activity in the dorsal and ventral parts of the premotor region was higher when viewing hand and mouth actions respectively (Buccino *et al.*, 2001; Aziz-Zadeh *et al.*, 2006; Wheaton *et al.*, 2004). Moreover, separate neural activities, specifically related to the observation of hand, mouth, and foot movements, were found in both frontal and parietal regions (Buccino *et al.*, 2001). Likewise, imaging studies showed that the left hemisphere exhibits a somatotopic arrangement along the motor strip (Schubotz *et al.*, 2003; Wheaton *et al.*, 2004; Leslie *et al.*, 2004) in response to action-related linguistic tasks, such as the passive reading of words (e.g. kick, pick, lick; Hauk *et al.*, 2004; Pulvermüller 2001), or phrases (Aziz-Zadeh *et al.*, 2006), or listening to sentences (Tettamanti *et al.*, 2005). High density EEG (Hauk *et al.*, 2006) and fMRI

(Gazzola *et al.*, 2006) investigations of healthy subjects revealed that the motor mapping of heard actions follows a somatotopic pattern. Indeed, clear differences in the topographies of brain responses to the sounds produced by finger and tongue clicking movements were found (Hauk *et al.* 2006). Hearing the sound associated with mouth actions and executing such action lead to activation of the pars opercularis of the inferior frontal gyrus, which extends to the rostral-most sector of the ventral premotor cortex. In contrast, while the hearing and execution of hand actions lead to activation of the ventral premotor cortex, this activation partially extends to parsopercularis of the inferior frontal gyrus (Gazzola *et al.,* 2006). In the left premotor cortex, a somatotopic pattern of activation was also observed with a dorsal cluster more involved when hearing and executing hand actions, and with a ventral cluster more involved when hearing and executing mouth actions (Gazzola *et al.*, 2006). These behavioural and neural double-dissociations indicate that a left-lateralized audio-motor network is actively involved in both the somatotopic motor mapping of sounds related to limb and mouth actions as well as in the execution of the very same actions. Note that the mouth sound region turned out to be represented ventrally to the hand sound area, in line with the well-known arrangement of the motor homunculus (Penfield and Rasmussen 1950) (Figure 1.1). This finding is in accordance with the topographical representation of different effectors within the ventral premotor, supplementary motor (SMA), and the primary motor (M1) cortex in monkeys (Takada *et al.*, 1998; Tokuno *et al.*, 1997).

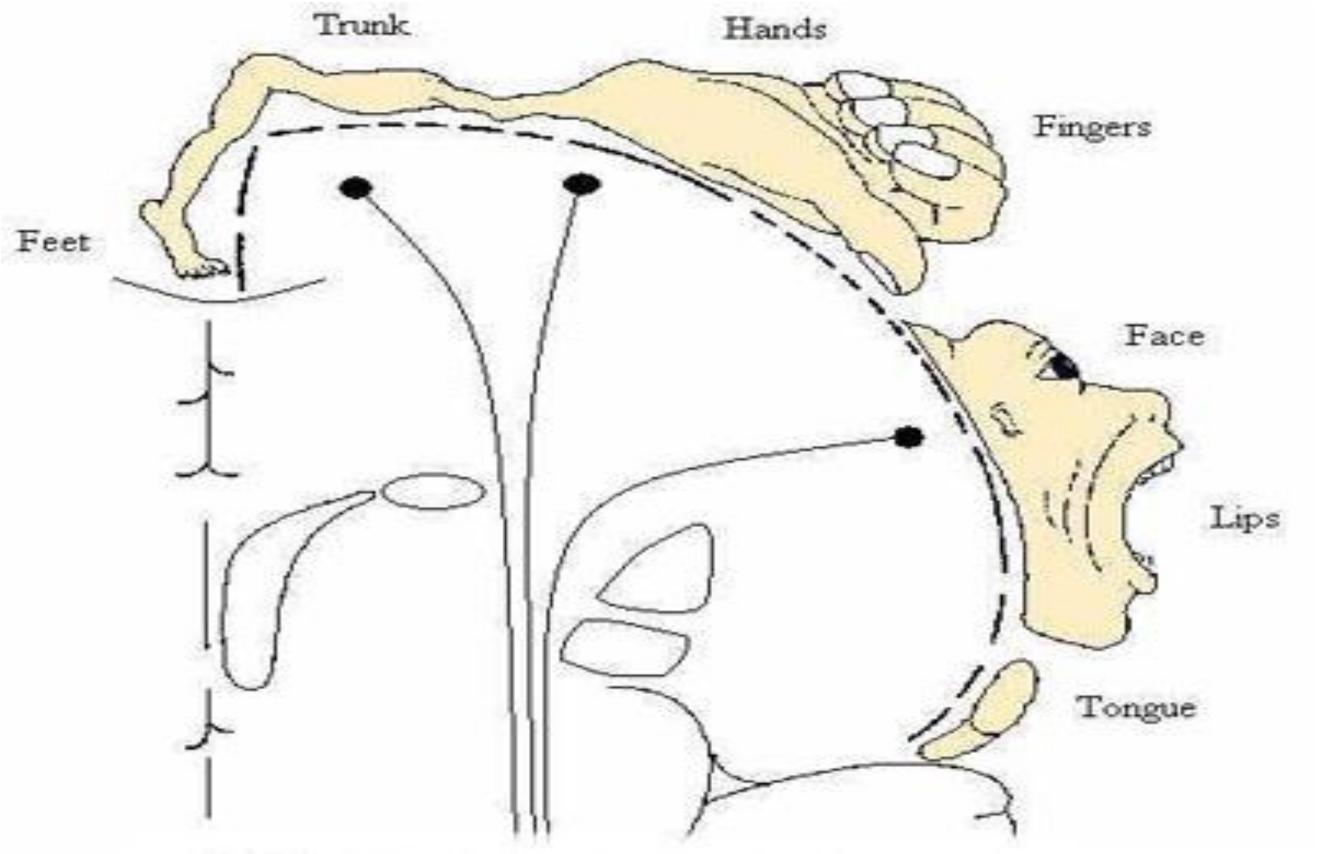


Figure 1.1: Somatotopic organization of the motor cortex

The existence of auditory mirror neurons suggest the view that the auditory-motor link does not depend on a previously established visual-motor link. Instead, would exist a direct connection between acoustic and motor representations. For example, it is noteworthy that in the Hauk and coworker's study (2006), the tongue sound, for which the visual information is irrelevant because the action is uniquely performed inside the mouth, elicited coherent somatotopic cerebral responses. For the aim of the next chapters this work is particularly interesting especially for the nature of the component studied (MMN). However, before to build the hypotheses underlie the next chapters I will rapidly review the main features of the MMN and its temporal and functional relation with the stream segregation phenomenon.

Chapter II

Deviance detection and Auditory Scene Analysis (ASA)

MMN AS AN INDEX OF DISCRIMINATION ACCURACY

The MMN is generated by the brain's automatic response to any change in auditory stimulation exceeding a certain limit roughly corresponding to the behavioral discrimination threshold (Tiitinen *et al.*, 1994). The MMN response is seen as a negative displacement in particular at the fronto-central and central scalp electrodes (relative to a mastoid or nose reference electrode) in the difference wave obtained by subtracting the event-related potential (ERP) to frequent, "standard", stimuli from that to rare "deviant" stimuli. The MMN usually peaks at 150-250 ms from change onset. A prerequisite of MMN elicitation is that the central auditory system has, before the occurrence of the deviant stimulus, been able to form a representation of the repetitive aspects of auditory stimulation (Winkler *et al.*, 1996a,b; Horváth *et al.*, 2001; see also Winkler *et al.*, 1999a,b; Huotilainen *et al.*, 1993; Paavilainen *et al.*, 1993a; for a review, see Näätänen and Winkler, 1999). An MMN is then elicited by a stimulus that violates this representation. Very importantly, in particular in view of the clinical and other potential applications, the MMN is elicited irrespective of the subject or patient's direction of attention (Näätänen, 1979, 1985).

Usually, the MMN is considered a quite accurate index of a pre-attentive feature-specific code of stimulus change and, further, it would provide an objective index of the discrimination accuracy for the different acoustic feature dimensions. This is supported by the fact that, in general, the MMN sensitivity to small stimulus changes seems quite well to correspond to the behavioural discrimination thresholds. In addition, the MMN's main parameters (i.e. latency and amplitude) correlate with the magnitude of sound change (Tiitinen *et al.*, 1994) as well as with one's perception (detection accuracy and speed) of the sound change (Lang *et al.*, 1990; Amenedo and Escera, 2000; Pakarinen *et al.*, 2007). Typically the MMN has been recorded using the oddball paradigm, where infrequent (probability P = 10-20%) "deviant" sounds are interspersed within a stream of continually repeated "standard" (P = 80-90%) sounds. Subsequent research has, however, shown that a single repeating constant standard sound is not a necessity for MMN recording (and elicitation). For instance, Huotilainen *et al.* (1993) recorded an MMN for a frequency change in a stream of sounds which constantly varied in intensity, duration, rise and fall times, and number of harmonic partials. Moreover, in the recent multi-feature recording paradigms (Näätänen *et al.*, 2004; Pakarinen *et al.*, 2007), the standard tone (P = 50%) is never repeated as it alternates with several types of deviants (for example P = 12.5% for each, 50% for all deviants in total). The multi-feature paradigms rest on the assumption that the deviants, which differ from the standard in one respect (one deviant differs in, e.g., duration, while another differs in frequency, yet others in loudness, vowel length, vowel duration, etc.) strengthen the memory trace for the standard in regard to those attributes they have in common. Crucially, has been demonstrated (Pakarinen *et al.*, 2010) that is possible to record an MMN even with no standard tone. This type of paradigm requires a sufficient number of independent deviant types in order for each of the deviant features to occur rarely enough (for example 12.5% for each deviant type). If one deviant type occurred with a large probability, it would no longer be a ‘‘deviant” event. When standard tones are not used, the other deviants take the role of the ‘‘standard” sound: they will strengthen the memory trace of one feature, e.g., they all share the same frequency, while the frequency deviant sound, occurring in 10 – 15% of cases, will act as a ‘‘deviant” for that feature. This could be possible for several other sound features, respectively. Theoretically, this work has very important implications. In fact, the MMN responses elicited in this no-standard paradigm, show concretely that the auditory system can, at the same time, keep track of different sound features by forming separated memory traces. In fact, the tones themselves could not be classified as standards or deviants in the traditional sense, as they all differed from each other, and were equiprobable. The different sound features, however, could be classified as common and rare, as their probabilities varied in the sequence. For instance, the frequency of the tones was identical in 87.5% of the trials, and either higher or lower in the remaining 12.5%. The elicitation of the MMN to frequency change under such conditions means that an accurate memory trace was constructed for the invariant (standard) features of the auditory input.

To deeply understand this very high and impressive ability to keep track of different acoustic features in trains of continuously changing sounds, several works has been focused on the neurophysiological basis promoting this accuracy. For example, Molholm and colleagues (2005), using functional magnetic resonance imaging (fMRI), showed that anatomically distinct networks of auditory cortices were activated as a function of the deviating acoustic feature (in this study tone frequency and tone duration). This work supported the neurophysiological plausibility that MMN generators in auditory cortex are feature-dependent. In this field, another important and pioneer magnetoencephalographic (MEG) work has been carried out by Tiitinen and coworkers (1993). Here the authors presented two tone stimuli, one frequent (standard) and the other one infrequent (a slightly higher, deviant tone) in random order to subjects reading texts they had selected. In different blocks, standards were either 250, 1000, or, 4000 Hz, with the deviant always being 10% higher in frequency than the standards of the same blocks. Magnetic responses elicited by the standard and deviant tones included N1m, the magnetoencephalographic equivalent of the electrical N1 (its supratemporal component). In addition, deviant stimuli elicited MMNm, the magnetic equivalent of the electrical mismatch negativity, MMN. The equivalent dipole sources of the two responses were located in supratemporal auditory cortex, with the MMNm source being anterior to that of N1m. The dipole orientations of both sources in the sagittal plane depended on stimulus frequency, suggesting that the responses were generated by tonotopically organized neuronal population (Figure 2.1).

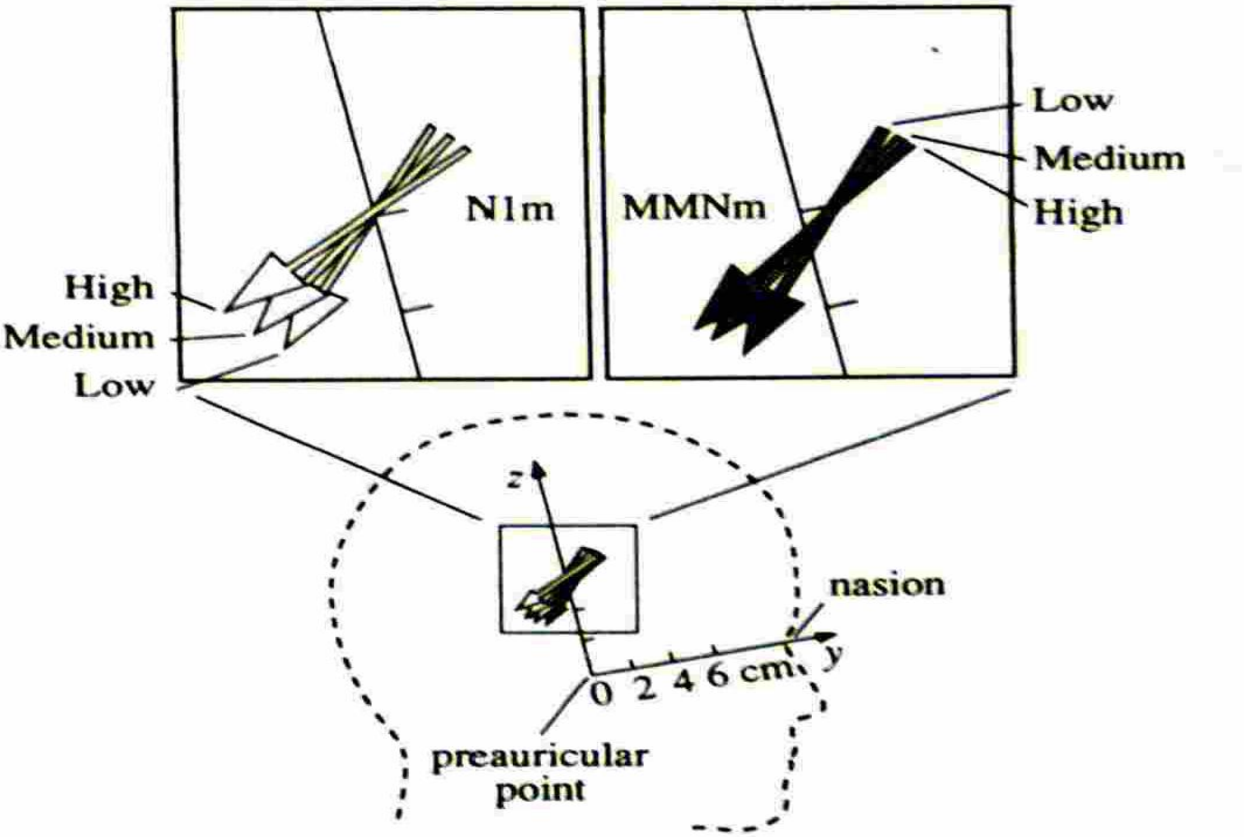


Figure 2.1: Mean angles of the N1m and MMNm dipoles in the sagittal plane for the low, medium and high tones.

The authors concluded stating that the tonotopy reflected by the frequency dependence of the MMNm source might be that of the neural trace system underlying frequency representation of auditory stimuli in sensory memory.

AUDITORY STREAM SEGREGATION

In daily life acoustic component generated by multiple sound sources often impinge upon the ear simultaneously. A primary task of the auditory system is to determine which elements in the acoustic mixture originate from which sound source, thereby constructing perceptual representations of the original sources. The ease with which the brain assigns sound components to their appropriate sources is illustrated, for example, at a cocktail party: speakers' voices, music etc. are perceived as distinct auditory object, despite the fact that the input to the ear is a complex sound wave arising from the summation of these acoustic signals. Auditory scene Analysis (ASA) is the process by which the auditory system groups and segregates components of acoustic mixtures to construct perceptual representations of sound sources, or ' auditory images' (Bregman, 1990). These auditory images in turn reflect the brain's determinations of the individuality of the sources generating the auditory signals (Bregman, 1990; Fay, 1998). Auditory scene analysis can be divided into two inter-dependent classes of processes, dealing with the perceptual organization of simultaneously and sequentially occurring acoustic elements, respectively (Bregman, 1990). Many of these processes are considered automatic, or 'primitive', in that they are thought to be based upon lower level neurophysiological mechanisms not dependent on learning or attention (Bregman, 1990). Acoustic features utilized by the auditory system in sound source determination are analogous to cues utilized in visual Gestalt perception. For example, acoustic elements arising from different spatial locations, or that are far apart in frequency or time, tend in nature to be generated by different sources and are perceptually segregated by the brain; sound component that are harmonically related or that rise and fall in intensity together (i.e. are co-modulated) tend to arise from a single source and are perceptually grouped. It has been maintained that scene analysis is the essence of hearing (Bregman, 1990; Yost, 1991; Fay, 1998). This assertion rests on the assumption that the world consists of distinct physical objects and events whose perceptual reconstruction from the complex flux of sensory input would clearly be of adaptive value to all organisms (Bregman, 1990).

A classic psychoacoustic phenomenon reflecting sequential organization in auditory scene analysis is called ‘auditory stream segregation’. This phenomenon is illustrated by listening to a sequence of temporally non-overlapping high and low frequency tones in an alternating pattern, ABAB. When the frequency separation (ΔF) between the tones is small (< 10 %), or their presentation rate is slow (< 10 Hz), listeners perceive a connected and coherent alternating sequence of high and low tones (i.e. galloping rhythm or coherence, see Figure 2.2). In contrast, when the ΔF is large or the PR is fast, coherence is lost and the alternating sequence perceptually splits into two parallel auditory streams, one composed of interrupted ‘A’ tones, and the other of interrupted ‘B’ tones, each perceived at half the PR (Miller and Heise, 1950; Bregman and Campbell, 1971; van Noorden, 1975; Anstis and Saida, 1985).

Noteworthy, it has been demonstrated that is Neurophysiologically plausible that the neural basis of this mechanism is based on the tonotopic structure of the auditory system. At this regard, most neurons in the auditory system, from the AN upwards, are frequency selective. This simple fact suggests that some aspects of streaming might arise from quite basic processes, such as could be observed at any neural site where this frequency selectivity is observed. Consider the response of a neuron tuned to the ‘A’ frequency , when ΔF is small, that neuron will also respond to the ‘B’ tones, and its output will reflect the ‘galloping rhythm’ in the sequence. As ΔF is increased, the neuron will respond only to the ‘A’ tone. There will be other neurons that only respond to the ‘B’ tones, but very few will respond to both. Now, when ΔF is intermediate, some neurons respond strongly to the ‘A’ tones and weakly to the ‘B’ tones. As the sequence is speeded up, the tones get closer together in time, and we might expect the short-term adaptation produced by the strong ‘A’ response to reduce the response to the ‘B’ tones. These findings have been observed in the primary auditory cortex (A1) of awake macaques (Fishman, Y.I. *et al.* 2001; Micheyl, C. *et al.* 2003).

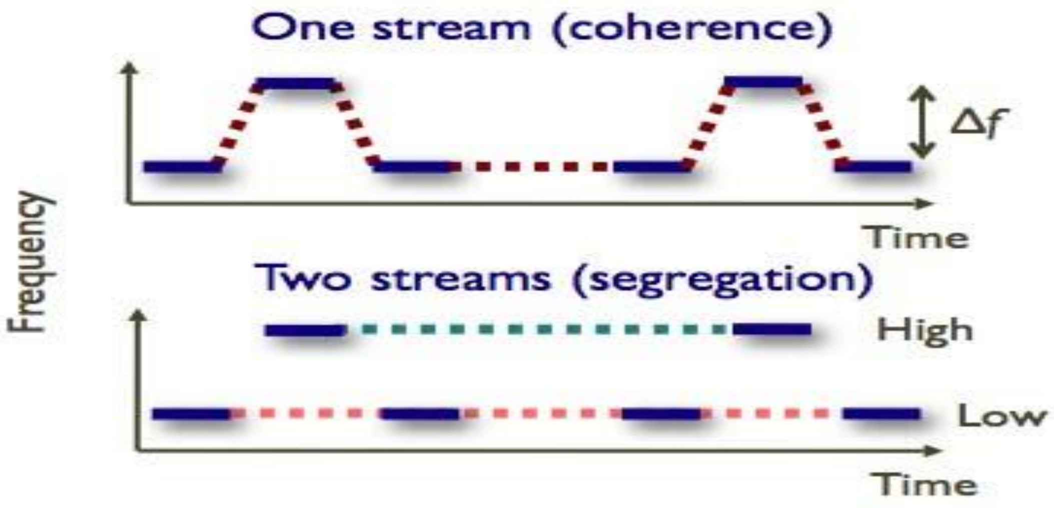


Figure 2.2: Stream segregation paradigm.

STREAM SEGREGATION AND MMN

Several studies convincingly demonstrated that stream segregation precedes, and provides a prerequisite for, MMN elicitation (Muller et al 2005b). Hence, the MMN is closely associated with the way the central auditory system organizes the incoming sounds: the sound organization determines the regularity on which MMN elicitation is based (see also Sussman *et al.*, 1998a; Winkler *et al.*, 2001; Alain *et al.*, 2001, 2002 for reviews, see Alain *et al.*, 1994; Näätänen *et al.*, 2001; Näätänen and Winkler, 1999).

CONCLUSION

The main goals of the works presented in the next chapters are based on the scientific background just delineated during this introduction. In particular, we will connect the evidences above exposed in order to propose, and test, a new approach to study the role of action representation in auditory scene analysis. In particular, this approach is based both on the strong evidences suggesting a feature-selective ability in acoustic representation and on the neurophysiological evidences above exposed. Indeed, this attempt is directly orientated to isolate the abstract feature “Motor Amenability” or “Motor content” of class of stimuli in the process of discrimination among sounds. At the basis of this hypothesis there are the previous data on MMN and action related sounds (Hauk 2006; and Lapage 2010) that indicated the role of action relatedness on sound discriminations. Furthermore, the analogous cortical organization between this systems (i.e. tonotopy vs somatotopy) provides the neurophysiological plausibility of this approach.

Chapter III

Investigating the stream segregation in audio-motor mapping

3.1 Introduction

Humans typically live in socially complex environments where elementary single movements are combined in complex and meaningful behavior which subtend heterogeneous intentions. All these actions usually produce characteristic sounds whose encoding is crucial to understand most of the daily behaviors. Different cognitive neuroscience techniques have provided convergent evidence for the existence of a fronto-parietal network involved in audio-motor mapping of human actions. Although, the audio-motor coupling is modulated by several factors, the body part involved in the action evoked by the heard sound (Fadiga *et al.* 2002, Pizzamiglio *et al.* 2005; Hauk *et al.* 2006; Gazzola *et al.* 2006; Pazzaglia *et al.* 2008a; Galati *et al.* 2008) seems play a prominent role. At this regard, several works reported coherent somatotopic activations following the perception of action-related sounds (for a review see Aglioti & Pazzaglia 2010). One strategy to deeply understand this resonance-like mechanism is to focus on the perceptual mechanisms mediated by analogous cortical organizations in different brain areas. At this regards, has been demonstrated that the perceptual discrimination and organization of consecutive sound stimuli (e.g. pure tones) are non-independent phenomena, partially based on the tonotopic organization of auditory cortex. This process provide the prerequisite for the Mismatch Negativity (MMN), a component of the event related potentials associated with pre-attentive sound discrimination. Several works demonstrated that the perceptual organization of successive sounds strongly determine the MMN elicitation and modulation (Muller *et al.*, 2005; Sussman *et al.*, 1998). Furthermore, has been demonstrated that both phenomena are based on the tonotopic organization of the auditory cortex (Tiitinen et al, 1993; Fishman *et al.*, 2001). Taking advantage of these evidences, we sought to determine whether the sound into action translation process is subjected to an analogous mechanism during sound discrimination. In particular, as the tonotopic structure of auditory cortex would play a role in both the discrimination and organization of successive sound stimuli by building arbitrary associations between them, we hypothesize that the somatotopic organization would play a very similar role in assigning arbitrary associations between sounds with different degree of motor amenability. To this end we studied the MMN. This component is usually recorded using the oddball paradigm, where infrequent (probability P = 10 - 20%) ‘deviant’ sounds are interspersed within a stream of continually repeated ‘standard’ (P = 80 - 90%) sounds. Several research have shown that a single repeating constant standard sound is not a necessity for the MMN elicitation (Huotilainen *et al.* 1993). For example, in the recent no-standard recording paradigm (Pakarinen *et al.*, 2010), the standard tones are not used, the other deviants take the role of the ‘‘standard” sound: they will strengthen the memory trace of one feature, e.g., they all share the same frequency, while the frequency deviant sound, occurring in 10 - 15% of cases, will act as a ‘‘deviant” for that feature. Taking advantage of these new experimental evidences, we created trains of sounds with the intent to isolate the abstract “motor amenability” feature in the process of discrimination of sounds. To this end, the multideviant blocks were designed so that the changing of the main acoustic parameters were too frequent, or too rare, to elicit an acoustic MMN.

3.2 Experiment 1: “The role of sounds organization in action perception”

Materials and Methods

PARTICIPANTS

Ten participants (5 males; mean 24.6 ± 2.7, years, all right handed), with no history of neurological psychiatric, or hearing impairment gave informed consent to participate in this study. This study was approved by the Ethical Committee of the Santa Lucia Hospital, Rome, Italy and was performed in accordance with the Declaration of Helsinki.

STIMULI

The original sounds were 2 stimuli (44.1 kHz, 16-bit, stereophonic): the water drop sound was composed by a single peak (400 ms), while the hand clapping was composed by two sounds in succession (460 ms). After have normalized the amplitudes, each original sound (D_O) has been used to create two multideviant blocks. One block was composed by the original sound and five stimuli created manipulating the D_O. The second block was composed by the same five sounds and by a disguised version of the original (D_{OD}). To this end, the peak frequency of a Fast Fourier Transform analysis was determined for each of the original stimuli (~ 970 Hz for the hand clapping sound and ~ 930 Hz for the water drop sound) (Figure 3.1). The stimuli were created, using the Cool

Edit 96 software (Syntrillium Software Corporation, Phoenix, AZ, USA) and Audacity 1.2.6 software, as follows (Figure 3.1).

Original disguised (D_{OD}): human and non-human original disguised sounds (D_{OD}) were generated by adding an inverse sine pure tone (3000 Hz) to the D_O, as long as it did not change the original peak frequency.

Standard ($S_{1:4}$): four stimuli were created by reducing the original peak frequencies of ~ 40% (~ 370 Hz for the hand clapping sound and ~ 365 Hz for the water drop sound) and by normalizing the amplitudes. After that, pure tones or noise at different frequencies and amplitudes were added as long as they did not alter the peak frequency: S_1 (square pure tone, 2500 Hz), S_2 (triangle pure tone, 2000 Hz), S_3 (sinusoidal pure tone, 1500 Hz), S_4 (white noise). In this way the $S_{1:4}$ had different frequencies distribution at lower amplitudes but the same peak frequency (Figure 3.1). Therefore, for this acoustic parameters they will act as standard sounds.

Low disguised (D_{LOW}): human and non-human D_{LOW} sounds were created by reducing the peak frequency of the original sounds of ~ 80% (~ 740 Hz for the hand clapping sound and ~ 730 Hz for the water drop sound) and by normalizing the amplitudes. These stimuli had the lowest peak frequency of the whole stimulus set (~ 230 Hz and ~ 200 Hz respectively).

Following this procedure the stimulus set is characterized by a continuous change of the main acoustic parameters among the stimuli. In particular, because the peak frequency and the signal strength are important parameters for the acoustic MMN elicitation, the peak frequencies and the Maximum Root Mean Squares (RMS) Power were controlled across all stimuli.

Peak frequency: within their respective multideviant block all the $S_{1:4}$ have the same peak frequency (~ 600 Hz for the hand clapping and ~ 565 Hz for the water drop sound). Therefore, the D_{O} and D_{OD} events represent always an increase of about 60% of their respective $S_{1:4}$ peak frequency (i.e. D_{O} and D_{OD} peak frequency: ~ 970 Hz for the hand clapping and ~ 930 Hz for the water drop). The D_{LOW} peak frequency events represent always a decrease of about 60% of their respective $S_{1:4}$ peak frequency (i.e. D_{LOW} peak frequency: ~ 230 Hz for the hand clapping and ~ 200 Hz for the water drop). Noteworthy, although the absolute difference between the S1:4 sounds and their respective $D_{O\text{-}OD}$ and D_{LOW} is different among the condition (i.e. hand clapping vs water drop), their relative amount of change is the same (i.e. about ± 60%). Finally, since we used different tones and noise at different pitches and amplitudes the resulting frequency distribution at lower amplitudes have a high level of variability among sounds (Figure 3.1).

Maximum Root Mean Square (RMS) Power: a 50 ms window width RMS revealed that, within their respective multideviant block all the $S_{1:4}$ have the same Maximum RMS Power ~ -14 dB for the hand clapping and ~ -8 dB for the water drop sound. The D_O and D_{OD} events with a Maximum RMS Power of ~ -16 dB for the hand clapping and ~ -11 dB for the water drop sounds, represent a very similar change in comparison to their respective $S_{1:4}$. Finally, the Maximum RMS Power of the D_{LOW} events was ~ -9 dB for the hand clapping and ~ -6 dB for the water drop sounds. Therefore, also for this parameters the $D_{O\text{-}OD}$ and D_{LOW} events represent a change and, hence, a further source of variability in the sequences of sounds. However, one important thing to take in consideration is also the variability of loudness in the sequence of sounds. Loudness is the characteristic of a sound that is primarily a psychological correlate of physical strength (amplitude). More formally, it is defined as "that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud" (American National Standards Institute, "American national psychoacoustical terminology" S3.20, 1973, American Standards Association). One important thing is that loudness is also affected by parameters other than sound pressure, in particular frequency and duration. Since all the sounds have the same length what we reduced was the possible influence that the frequency difference could have on this subjective dimension. At this regard, it is well know that the sensitivity of human hear changes also as a function of frequency as revealed by the equal-loudness contour (latest version ISO

226:2003 revision). In particular, the frequencies below 2 - 4 kHz needs progressively higher sound pressure level in order to elicit similar loudness level. Therefore, since our D_{LOW} events are characterized by a very low peak frequency they could elicit an MMN based on this subjective dimension. To avoid this scenario the S_4 has been created with the intent to further increase the variability of the average intensity. Therefore, the average RMS revealed a high level of variability among sounds (Average RMS: hand clapping D_O = -31 dB, D_{OD} = -32 dB, D_{LOW} = -27 dB, $S_{1:3}$ = -29 dB, S_4 = -22 dB; water drop D_O = -25 dB, D_{OD} = -26 dB, D_{LOW} = -20 dB, $S_{1:3}$ = -22 dB, S_4 = -16 dB).

Length and temporal evolution of RMS Power: within one multideviant block all the stimuli have the same length that is 460 and 400 ms for the hand clapping and water drop sound, respectively. Finally, because the latency of the acoustic peak signal strength can affect the MMN, we controlled also that the $D_{O\text{-}OD}$ and D_{LOW} events had the acoustic peak latency as similar as possible. This is important because the filters that we applied to create the D_{LOW} events could affect the sound by stretch it and, hence, shifting the local peaks. When this shifting is kept down it can slightly affect the MMN latency. However, if it was too big it could dramatically change the perception of the sounds by introducing a very different power distribution of the sound signal over time. Therefore, the Maximum RMS Power latencies for the hand clapping D_O and D_{OD} were both ~ 30 ms while their respective D_{LOW} stimulus showed its Maximum Power at 40 ms. The water drop D_O and D_{OD} events had their Maximum Power at ~100 ms while the water drop D_{LOW} event showed a latency of this measure at ~120 ms. Although, there is not a perfect matching between the $D_{O\text{-}OD}$ and D_{LOW} events respects the latency of the Maximum Power, this is very limited in comparison to the stimulus length and, most important, very similar among the hand clapping and water drop conditions.

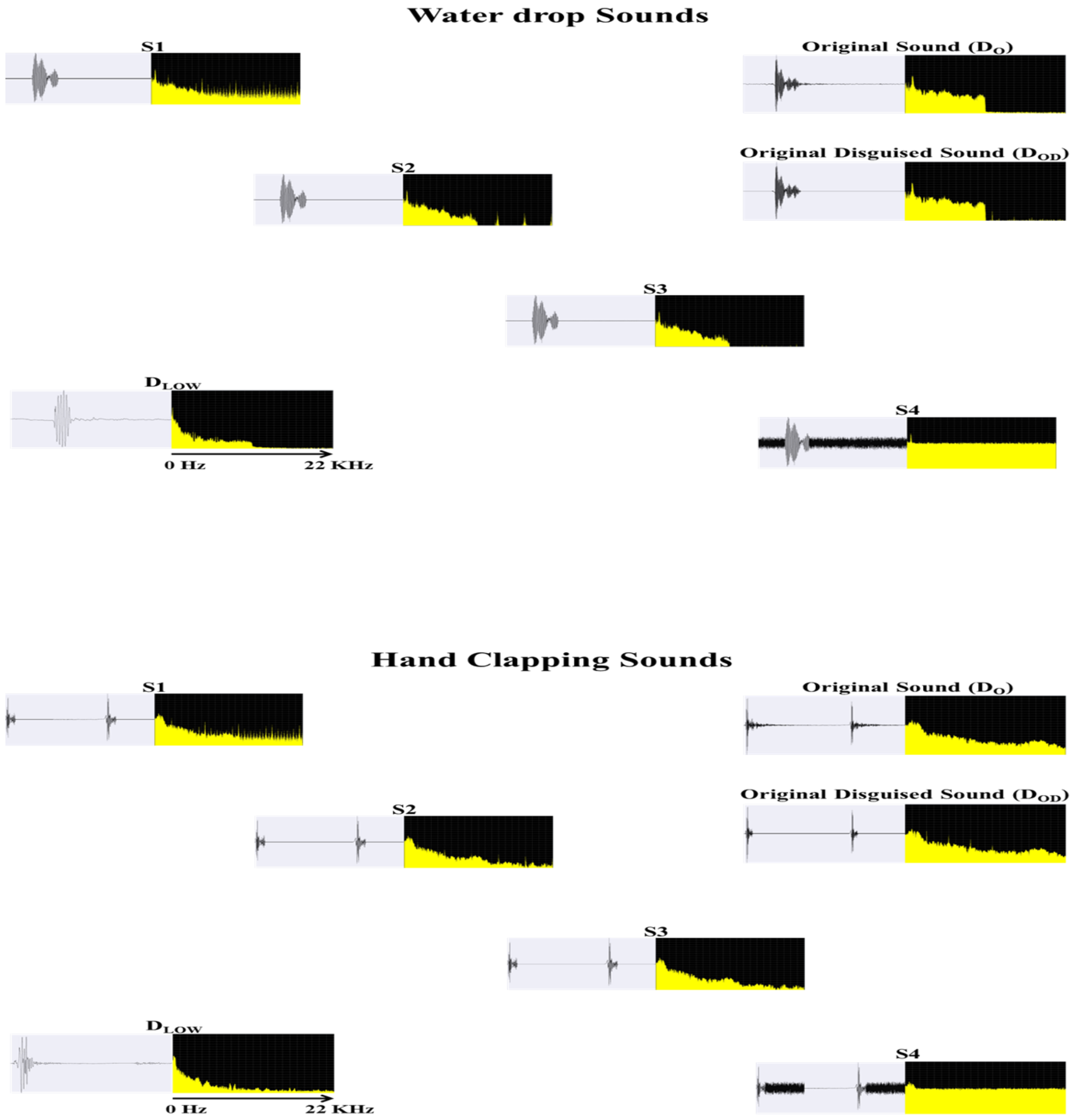


Figure 3.1: Description of acoustic stimuli. Per each stimulus category (water drop and hand clapping) and stimulus type ($S_{1:4}$, D_O, D_{OD} and D_{LOW}), the waveform sounds (left) and frequency power distributions (right) are plotted.

GENERAL PROCEDURE

To evaluate the subjective rating of the perceptual differences between sounds, a pre-test session was conducted. Participants were asked to seat in a comfortable chair and to assess the entire stimulus set, binaurally delivered via headphone in randomized order, in two separated visual analogue scales (VAS), ranging from 0 to 100, evaluating: (1) “How do you think this sound is reproducible by human body performing movement or action?” (2) “How frequently do you produce similar sounds performing movement or action?” (3) “How do you think this sound is reproducible by water? Or, How is it water-related to you? After that, the participants were instructed to concentrate on watching a documentary (“Microcosmos”, 1996), presented without audio, throughout the EEG recording and were told that the acoustic stimuli were of no relevance to them, so they had to pay attention only to the movie. Indeed, the participants listened to four different multideviant blocks. Two blocks were composed by each original human and non-human action related sounds (D_O) and by their respective $S_{1:4}$ and D_{LOW} events. Two further blocks were composed by the same $S_{1:4}$ and D_{LOW} events and by their respective D_{OD} events. The order of the Multideviant blocks was counterbalanced across subjects, with the restriction that the Hand clapping and Water drop sounds blocks were always alternated. The occurrence of the deviants was pseudo-randomized so that each deviant was presented once in an array of six successive deviants and that two successive deviants always were of different type. Each deviant with extreme peak frequency (Original, D_{OD} and D_{LOW}) was presented 110 times while all the other sounds 140 times. Therefore, the $S_{1:4}$' peak frequency had a higher probability (~70%) than the $D_{O\text{-}OD}$ and D_{LOW} (~15% per each stimulus). However, taken together, the event "change of the peak frequency" occurred with a probability too high (~28%) to elicit an acoustic MMN (Pakarinen *et al.*, 2010). The stimulus-onset-asynchrony (SOA) was 400 ms, the total recording time per each Multideviant block was always about 15 minutes. Finally, in order to obtain a difference signal we created the Repetition block in which we presented repetitively the Original, D_{OD} and D_{LOW} 110 times, while S_1 was presented 140 times as in the Multideviant block. Stimulus presentation, timing and pseudo-randomization were controlled by using E-Prime ver.1.2 software (Psychology Software Tools Inc., Pittsburgh, PA) running on a PC, XP operative system.

ELECTROPHYSIOLOGICAL RECORDING

EEG was recorded and amplified by Neuroscan SynAmps² System and by using 64 tin electrodes embedded in a fabric cap (Electro-Cap International), arranged according to the international 10-10 system. Horizontal bipolar electro-oculogram was recorded from electrodes placed on the outer canthus of each eye, and vertical electro-oculogram was recorded from an electrode below the right eye. All electrodes were referenced to the activity recorded at an electrode placed on the nose. For the whole acquisition the EEG data were digitized with a sampling rate of 1000 Hz, and with an amplifier filter band-pass DC−100Hz, the impedance for all electrodes was kept below 5 KΩ. Offline analyses were carried on with analyzer 1.05 (Brain products GmbH). The sampling rate was reduced to 250 Hz and then filtered using a digital 1 Hz high-pass filter. The epochs for the stimulus-locked ERPs were 600ms, including 100ms pre-stimulus baseline. After data segmentation, Independent Component Analysis (ICA) with standard parameters for artifact removal as implemented in EEGLAB 10 (Swartz Center for Computational Neurosciences, La Jolla, CA; http://www.sccn.ucsd.edu/eeglab) was performed on the basis of the following criteria: a component was considered to be artifactual when its topography showed peak activity only over the horizontal or vertical eye electrodes and when it showed a smoothly decreasing power spectrum (which is typical for eye movement artifacts, see Delorme and Makeig, 2004). After calculating the independent components, eye blink and eye movement components were subtracted from the EEG data. On average 3 (range 2 – 6) components were removed from each subject. After that, the data were filtered using a digital 20 Hz low-pass filter. Epochs with voltage fluctuation > 80 µV in VOEG channel and > 50 µV in HEOG channel and those contaminated with artifacts due to amplifier clipping, bursts of electromyographic activity, or peak-to-peak deflection exceeding ±80 µV were excluded from averaging by a semi-automatic rejection procedure. On average approximately 5% (range 3 – 6%) of the trials were rejected because they violated these artifact criteria. The MMN was computed by subtracting ERP responses of the sounds presented in the repetition block (Original, D_{OD}, D_{LOW} and S_1) from ERP responses elicited by the same stimuli presented as deviants in the multideviant block (Näätänen *et al.*, 2004; Pakarinen *et al.*, 2010). The peak of the grand-average response was separately determined for all difference signals in the 30–280ms time window. The mean amplitudes were defined as the mean voltage of a 40 ms time window (± 20 ms) centered at the peak latency at FCz in the grand-average difference signal.

DATA ANALYSIS

Potential effects of sound processing were assessed for MMN amplitudes at channel FCz referenced to the electrode placed on the nose, in a $2 \times 2 \times 2$ repeated measures ANOVA with the following factors: Context (i.e., presence of D_O, D_{OD}), Sound Category (human, non-human), Deviant (D_O-D_{OD}, D_{LOW}). Furthermore, in order to verify that the Standard events did not elicit any MMN we carried out a t-Test against zero on the S_1 mean amplitude.

Statistical analysis of the subjective ratings was carried out using three separate 2 x 7 Repeated Measures ANOVAs, one for each rating dimension. The factors were ‘Sound’ (Hand Clapping, Water Drop) and ‘Deviant’ (D_O, D_{OD}, D_{LOW}, $S_{1:4}$). All pairwise comparisons were performed using the Newman–Keuls post-hoc test. Partial eta-squared (ηp^2) was selected as the index of effect size: 0.01–0.06 = small effect, 0.06–0.14 = moderate effect, and > 0.14 = large effect (Cohen, 1973). A significance threshold of $p < 0.05$ was set for all statistical analyses. We used, when appropriate, the Greenhouse–Geisser correction for sphericity.

RESULTS

VISUAL ANALOGUE SCALE

Repeated measures ANOVAs of the three ratings for listened sound revealed main effects of Sound (all $F_{1,9} > 13.0$, $p < 0.006$, $\eta p^2 > 0.58$) and Deviant (all $F_{6,54} > 13.50$, $p < 0.001$, $\eta p^2 > 0.59$). However, since we are exploring the meaning dimension in the discriminative process among sounds this main effects are not very informative. In fact what we need to know are the relative differences between the deviants and standards. At this regard, the Sound × Deviant interaction (all $F_{6,54} > 5.40$, $p < 0.001$, $\eta p^2 > 0.37$) was observed (Figure 3.2). Newman–Keuls post-hoc test revealed that, for their respective rating dimension, the original sounds were significantly better perceived as sounds feasible as human or non-human actions than their respective D_{LOW}, $S_{1:4}$ and D_{OD} sounds (all $p_s < 0.004$). Importantly, for their respective rating dimension the D_{LOW} sounds, even if significantly less well perceived as feasible actions compared to their respective original sounds (all $p_s < 0.004$), they were significantly better perceived as feasible actions than their respective $S_{1:4}$ (all $p_s < 0.007$). This is coherent with the fact that D_{LOW} sounds were created by disguising less the original sound than all the $S_{1:4}$ events, and thus appear more distinctly natural as human or non-human feasible action sounds. Noteworthy, the D_{LOW} events were evaluated not statistically different between them in all three scales (all $p_s > 0.29$). Finally, in their respective rating dimension, no significant differences were found between $S_{1:4}$ and D_{OD} sounds (all $p_s > 0.17$).

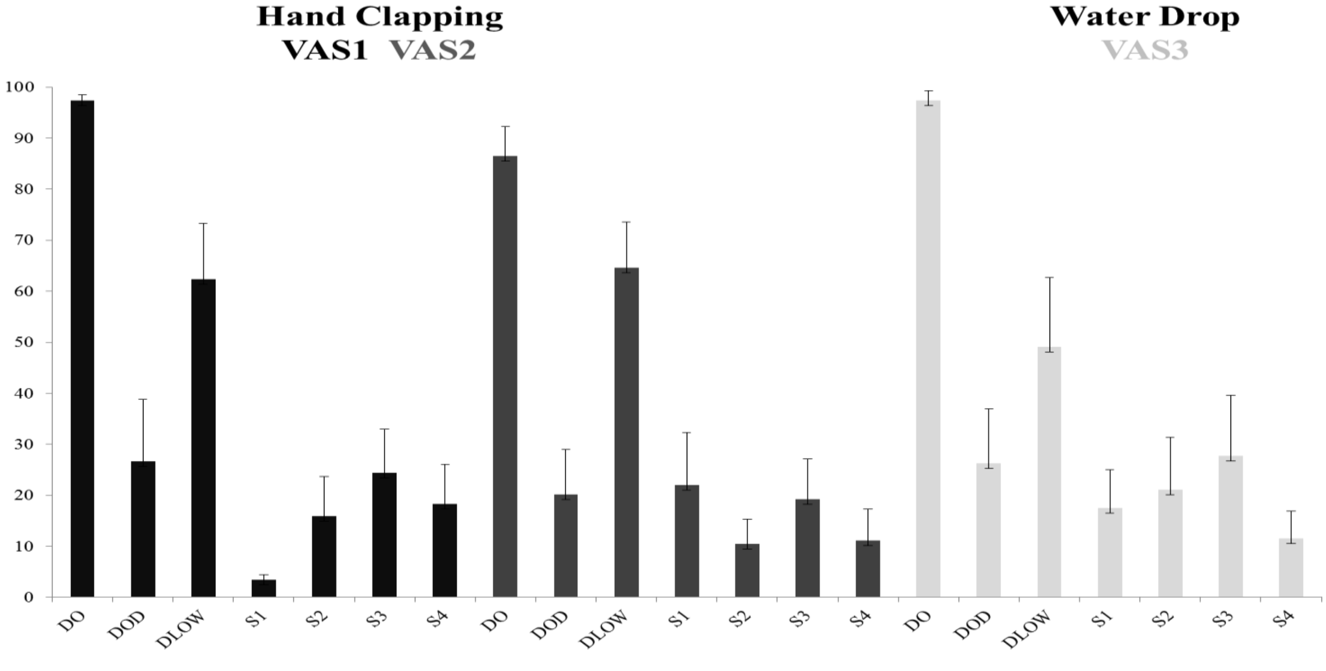


Figure 3.2: Mean + SEM ratings (scale 0-100) on their respective rating dimension. The first fourteen scores refers to the hand clapping sounds for their respective rating dimension (VAS 1 and 2, see above). The last seven bars refer to the water drop sounds for their respective rating dimension (VAS 3, see above).

MISMATCH NEGATIVITY

In the 30-280 ms time window, a negative ERP deflection was observed at fronto-central electrodes (Figure 3.3). The Repeated Measures ANOVA on the MMN mean amplitudes revealed significant main effects of Context and Sound (for both, $F_{1,9} > 12.9$, $p < 0.006$, $\eta p^2 > 0.59$). These effects show that the MMN amplitudes were significantly smaller in the presence of the D_{OD} than in the presence of the D_O sounds ($p < 0.001$), but also that the MMN amplitudes were modulated by the sound type. In particular, the hand clapping events were significantly bigger than the analogous water drop events ($p < 0.006$). Furthermore, the Context × Sound, and Context × Sound × Deviant interactions (for both, $F_{1,9} > 5.5$, $p < 0.05$, $\eta p^2 > 0.38$) indicate that the presence of the original sound affects particularly the hand clapping's deviants ($p < 0.001$), and in particular it has a differential effect on them. Newman-Keuls post-hoc test for multiple comparisons showed that the MMN amplitudes of the human D_{LOW}, presented with its Original sound, and the human D_O sound were significantly bigger than all the other MMN mean amplitudes entered in this ANOVA (all $p_s < 0.03$). Therefore, it is important to note that the MMN elicited by the hand clapping D_{LOW} sound, presented with its Original sound, was significantly bigger in comparison to the MMN elicited by the same sound when presented with the human D_{OD} ($p = 0.02$). All the other comparisons entered in this ANOVA were not significantly different among them (all $p_s > 0.38$). Finally, all the S_1 events did not elicit MMN responses (t-Test against zero, all $p_s > 0.12$).

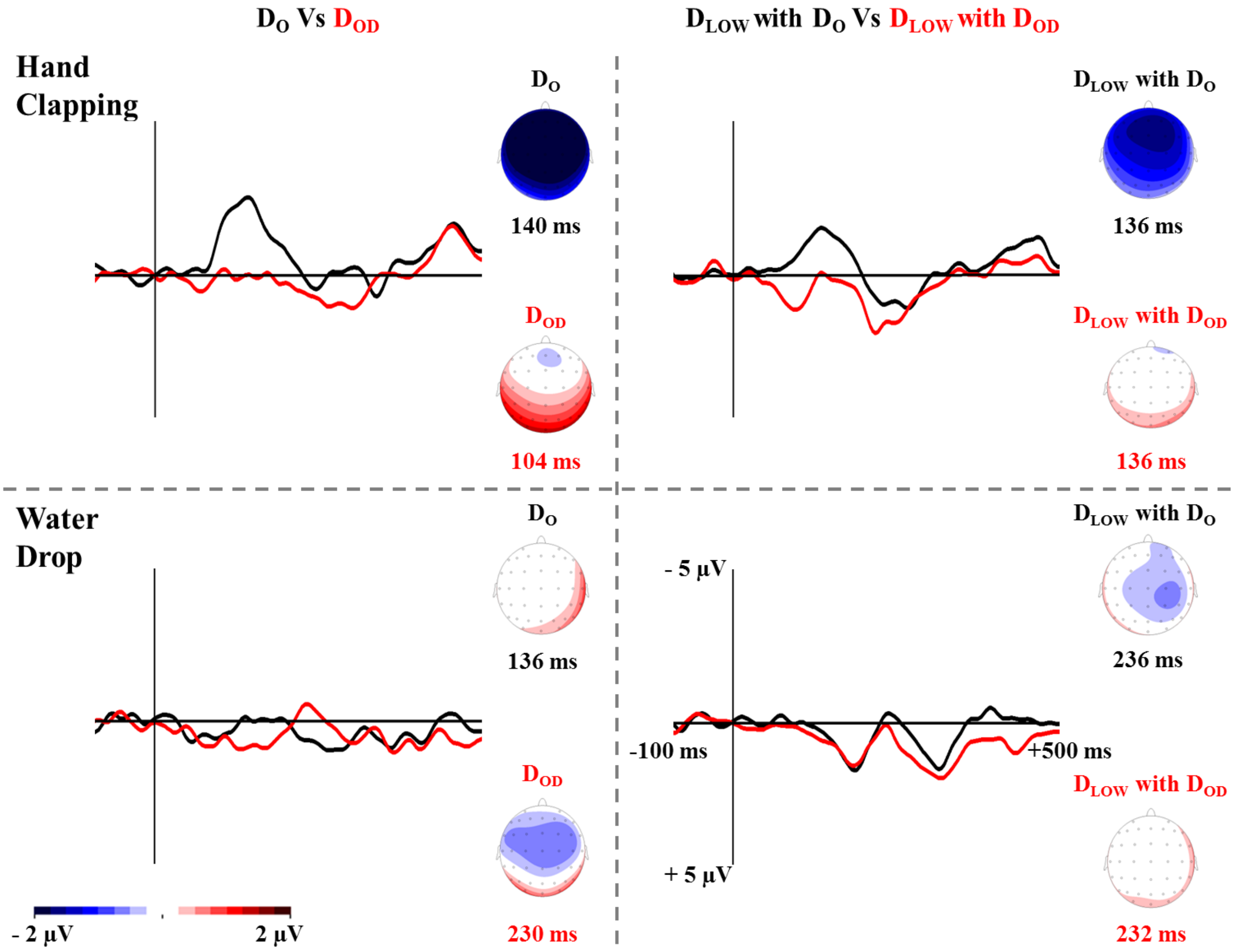


Figure 3.3: Grand average of nose-referenced deviant-minus-repeated difference signals of 10 subjects at FCz electrode for all sounds that entered in the final analysis.

Overall, this results indicate that the presence of intelligible human action sounds increase the likelihood that auditory-motor associations extend to less intelligible ones. This means that during sound discrimination the "motor amenability" dimension affect the way in which the stimuli are organized and, hence, discriminated. This doubleprocess of organization and discrimination has been extensively investigated in acoustic processing where has been demonstrated that the way in which the brain organize the percepts has a strong influence on sound discrimination (Muller *et al.*, 2005; Sussman *et al.*, 1998). Although at an abstract and subjective level, this result would suggest a very similar mechanism to the phenomenon of stream segregation. This analogy is supported by considering the modulation of the hand clapping D_{LOW} event depending on the level of action association of the deviant with the highest peak frequency (i.e. D_O vs D_{OD}). We speculate that the D_O and D_{LOW} sounds became fused within the same domain (i.e. motor amenability) but segregated within the specific motor association (e.g. coupled with different motor representations). Therefore, this sounds would be grouped and distinguished from the $S_{1:4}$ under the same general heuristic (i.e. motor amenability). However, the specific motor connotation must be different because otherwise they should not more be able to elicit an MMN because they would occur too frequently (i.e. 30%). Therefore, it is highly plausible that some form of segregation occurs at this very specific stage of processing. However, to directly test these hypotheses we carried out a second experiment in which the same paradigm has been applied to different human action related sounds whose pitches were different. Therefore, in order to have comparable MMN responses between Multideviant blocks, we have to change the discriminative acoustic feature (i.e. the peak frequency) by the same relative amount between the D_O and D_{LOW} events and their respective $S_{1:4}$. However, since the original pitches were different, the absolute changing between sounds was different. This would lead to different subjective distances between the D_O and D_{LOW} events in the different multideviant blocks.

3.3 Experiment 2: “The role of perceptual salience in audio-motor mapping”

Materials and Methods

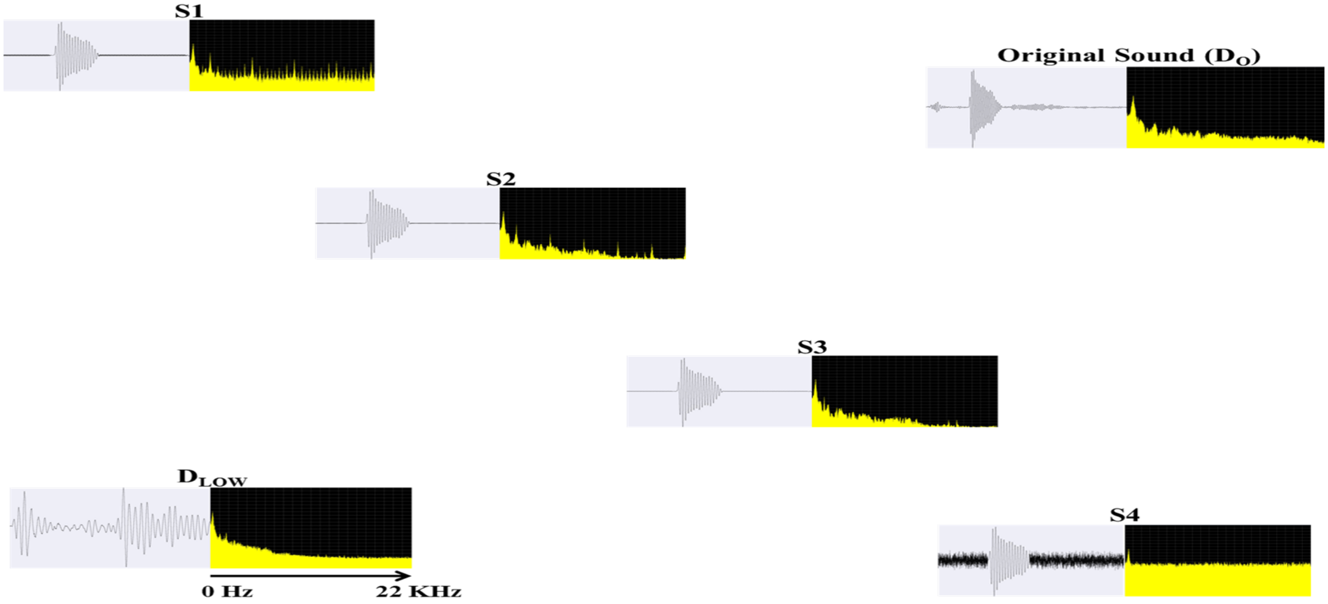
PARTICIPANTS

Ten participants (7 males; mean 35.9 ± 6.4 years, all right handed), with no history of neurological psychiatric or hearing impairment gave informed consent to participate in this study.

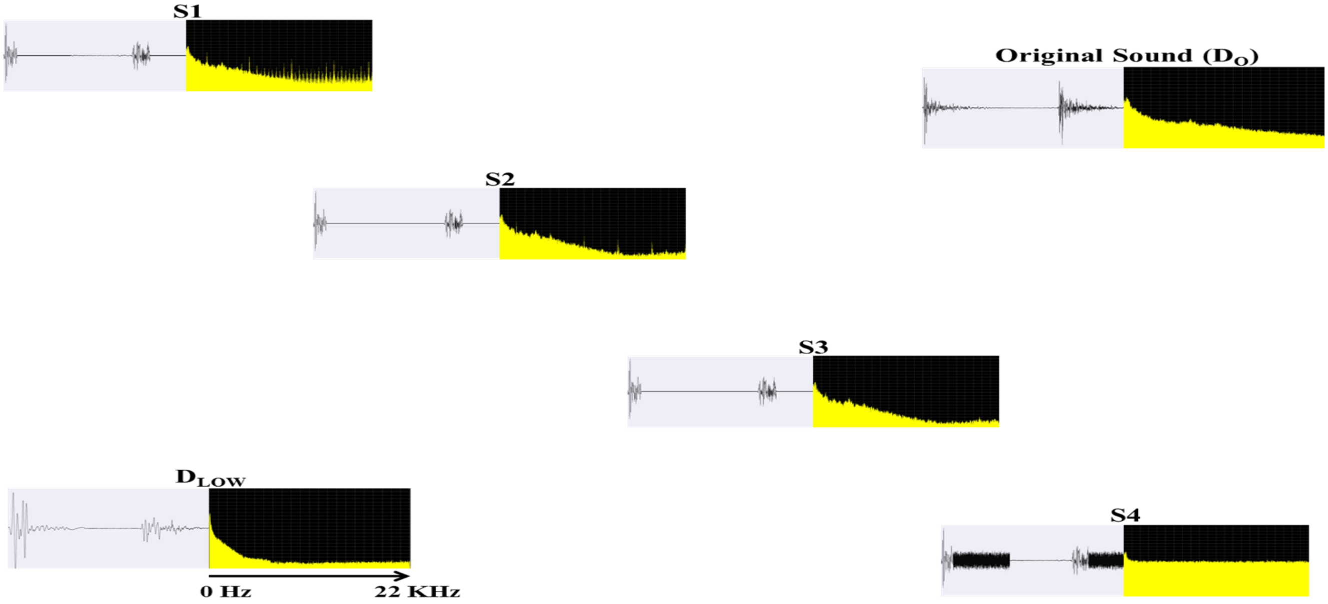
STIMULI AND PROCEDURE

Two more human action sounds (44.1 kHz, 16-bit, stereophonic), the tongue click sound, composed by a single peak (140 ms) and footstep, composed by two sounds in succession (700 ms) were tested. Each of human (hand clapping, tongue click, footstep) and non-human (water drop) action sound has been used to create one’s own multideviant block composed by the D_O, D_{LOW} and the $S_{1:4}$ events. The stimuli were created with the same procedure previously presented (see above). It is important to note that, in order to have comparable MMNs among paradigm, the acoustic feature that marks the D_O and D_{LOW} sounds (i.e. the peak frequency) must have the same relative amount of changing in comparison to their respective standard sounds (i.e. always ~ 60% in all multideviant blocks). However, because the different action-related sounds have different pitches, the absolute difference of this acoustic parameter is different among the sounds (Figure 3.4). In particular, the original footstep sound had the lowest peak frequency (~ 360 Hz) and, therefore, it has been less decreased to generate the $S_{1:4}$ and D_{LOW} (i.e. ~ 225 Hz and ~ 90 Hz of peak frequency, respectively). The tongue click sound, that was higher in pitch (~ 810 Hz), had a greater amount of absolute changing (i.e. peak frequency at 510 Hz for the $S_{1:4}$ and at 210 Hz for the D_{LOW}). Therefore, the deviants represent always a 60% of increase (D_O) or decrease (D_{LOW}) of the standards peak frequency. Following this procedure we have comparable acoustic changes between the multideviant blocks, but different subjective distances between the D_O and D_{LOW} events.

The experimental procedures, the subjective rating, the electrophysiological recording methods and the preprocessing stages were identical to those exposed in Experiment 1(see above). Each of the deviant with the extreme peak frequency (i.e. D_O and D_{LOW}) was presented 180 times, whereas the $S_{1:4}$ were presented 210 times. The SOA was constant at 500 ms. On average 2.7 (range 2 – 4) components were removed from each subject. Furthermore, approximately 5% (range 3 – 6%) of the trials were rejected because they violated the artifact criteria.



Tongue Click Sounds



Footstep Sounds

Figure 3.4: Description of acoustic stimuli. Per each new stimulus category (i.e. Footstep and Tongue Click) and stimulus type (S1:4, DO and DLOW), the waveform sounds (left) and frequency power distributions (right) are plotted. For the hand clapping and water drop stimuli see above.

STATISTICAL ANALYSIS

Potential effects of deviant sound processing were assessed for MMN amplitudes at channel FCz, referenced to the electrode placed on the nose, by 4 × 2 repeated measures ANOVA with the following factors: Sound (i.e., tongue click, hand clapping, footstep, and water drop) and Deviant (i.e., D_O and D_{LOW}). Furthermore, in order to test topographical differences between the Original and D_{LOW} events, the MMNs were extracted for each subject at electrodes F7, F3, Fz, F4, F8 / T7, C3, Cz, C4, T8 / P7, P3, Pz, P4 and P8. They were grouped into the factors Gradient (anterior-posterior, three levels) and Laterality (left-right, five levels), and subjected to a new ANOVA with the factor Sound (tongue click, hand clapping, footstep and water drop) and deviant (Original, D_{LOW}).

The subjective ratings of auditory familiarity and perceived motion in the human and non-human action sounds were compared using three separate repeated measures ANOVAs. The factors were Sound (tongue click, hand clapping, footstep and water drop) and Deviant (D_O, D_{LOW} and D_{HIGH}1:4).

SOURCE ESTIMATION

The source estimation has been carried out on unsubtracted ERP using the standard procedure implemented in SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/). This approach results in a spatial projection of sensor data into (3D) brain space and considers brain activity as comprising a very large number of dipolar sources spread over the cortical sheet, with fixed locations and orientations. This renders the observation model linear, the unknown variables being the source amplitudes. This standard procedure is divided into four consecutive steps, which characterized any inverse procedure with an additional step of summarizing the results. The first three steps (source space modeling, data coregistration and forward computation) specify the forward or generative model. The fourth step (i.e. inverse reconstruction) is concerned with Bayesian inversion of that model. The reconstructed activity in 3D voxel space has been analyzed using mass univariate analysis in SPM, using appropriate summary statistic images over the same 40 ms time window used to analyze the MMN mean amplitude (see above). Therefore, the statistical maps were tresholded at $p < 0.001$ and further corrected for multiple comparison (FDR correction at cluster level).

RESULTS

VISUAL ANALOGUE SCALE

The main effects of Sound (all $F_{3,27} > 8.51$, $p < 0.001$, $\eta p^2 > 0.48$) and Deviant (all $F_{5,45} > 6.62$, $p < 0.001$, $\eta p^2 > 0.41$) indicated that, for the respective rating dimension, there was a modulation of the perceived motion as human or non-human action across sounds and kind of disguise. Furthermore, the interactions Sound × Deviant ($F_{15,135} > 2.53$, $p < 0.003$, $\eta p^2 > 0.21$) were observed in all the rating dimensions (Figure 3.5). Newman-Keuls post-hoc test revealed that in both the visual analogue scales the Tongue Click Original sound was the most immediately recognizable compared to all its sounds (all ps < 0.001), which were similarly recognized among them (all $p_s > 0.07$). Here we have the maximum distance between the original and the D_{LOW}. Similarly, in both the visual analogue scales the hand clapping original sound was more recognizable in comparison to its D_{LOW} (both $p_s < 0.001$) and to its $S_{1:4}$ (all $p_s < 0.001$) events. The evaluation of the hand clapping D_{LOW} event in the visual analogue scale assessing the general motor amenability of the sound (VAS 1 see above) revealed that this sound was better perceived as human feasible action sound in comparison to all the $S_{1:4}$ (all ps < 0.02). The visual analogue scale assessing the probability to produce this sound performing movement in daily life (VAS 2 see above) revealed a very similar, even if smaller, pattern of result (all $p_s < 0.06$). Finally, in both scales the hand clapping $S_{1:4}$ did not show any significant difference among them (all ps > 0.48). Here, we have an intermediate distance between the D_O and D_{LOW} events on the 'motor amenability' dimension. In the VAS 1 the footstep original sound was not more significantly recognizable than its D_{LOW} ($p > 0.09$), and both of them had significant higher scores than all the other deviants (all ps < 0.001) which were similar among them (all $p_s > 0.84$). However, the VAS 2 revealed that only the original footstep sound was more recognizable than all the other D_{LOW} and $S_{1:4}$ events ($p < 0.001$) and this is coherent if we assume that when we directly ask about the daily life of participants they became more conservative. Therefore, coherently with the absolute amount of peak frequency reduction, we have that the general motor amenability scale (VAS 1) revealed a short distance between the original and the D_{LOW}. Finally, within its rating dimension the water drop original sound was the most recognizable compared to all its deviants ($p < 0.001$), while the D_{LOW} and all the other sounds had VAS scores not significantly different between them (all $ps > 0.41$). Here, as the tongue click sound, we have the maximum distance between the original and the D_{LOW}. Overall, this results seem reasonably indicate that the absolute amount of peak frequency reduction strongly affects the way in which the sounds are perceived.

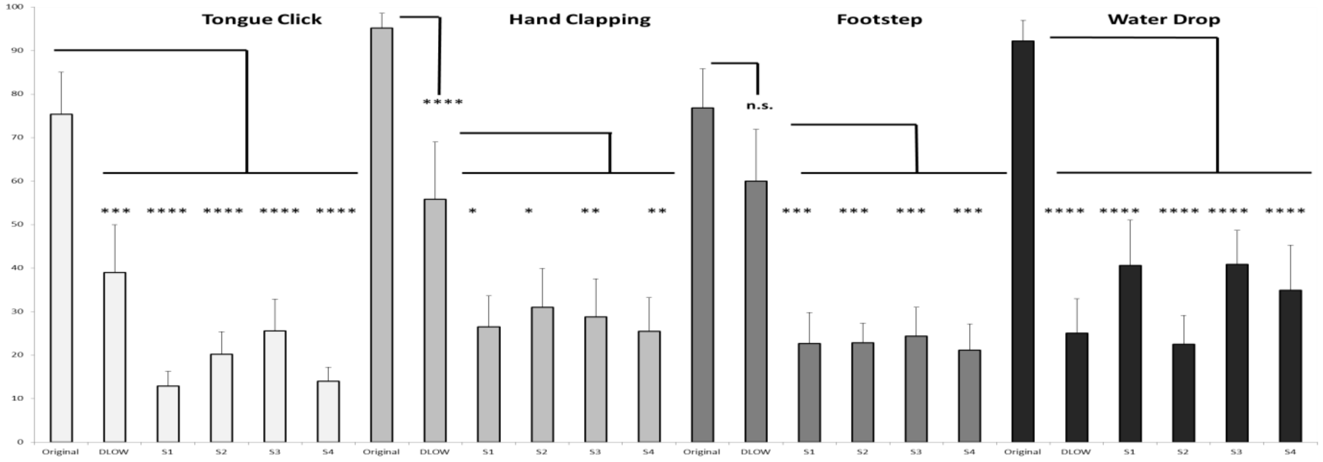


Figure 3.5 Mean + SEM ratings (scale 0–100) on relevant stimulus dimensions Action- and Water-relatedness respectively (VAS 1 and 3).

MISMATCH NEGATIVITY

In the 30-280 ms time window, a negative ERP deflection was observed at fronto-central electrodes (Figure 3.6). The repeated measures ANOVAs on the MMN mean amplitudes revealed a main effect of Sound ($F_{3,27} = 26.79$, $p < 0.001$, $\eta p^2 = 0.75$). Newman-Keuls post-hoc test for multiple comparisons revealed that the tongue click and hand clapping sounds (D_O and D_{LOW} collapsed) were not significantly different between them ($p = 0.99$). However, both this sounds elicited a bigger MMN in comparison to the footstep (both $p_s < 0.004$) and water drop sound (both $p_s < 0.001$). Crucially, the footstep sound elicited an MMN significantly bigger in comparison to the water drop events ($p < 0.001$). Differently to the previous experiment we did not observe the interaction of the sound and deviant factor. One possible explanation is that in this experiment we presented more stimuli (i.e. 180 repetition, see above). This could allow a sort of strengthen of the association between the D_O and D_{LOW} mediated by learning process. Therefore, in order to compare the MMN mean amplitude between sounds and deviants we carried out planned comparisons. When we contrasted the tongue click and hand clapping D_O and D_{LOW} events we did not observe any significant differences (both $p_s > 0.45$). This result is compatible with the subjective ratings where the original sounds were well distinguished in comparison to their respective D_{LOW} events. Therefore, we speculate that, although they were fused under the same generic motor classification they were kept segregated within this dimension. This hypothesis seems supported by considering that when we contrasted both deviants of this sounds against the footstep D_O and D_{LOW} events we observed a strong tendency to differ (all $p_s < 0.053$). Again, consistent with the absolute degree of masking of sounds and with the subjective ratings is highly probable that the footstep sounds were more fused to each other than the tongue click and hand clapping. Therefore, this modulation of the MMN is coherent with a phenomenon of stream segregation (i.e. more close within one stream two sounds are, more difficult is their discrimination and lower would be their MMN). At this regard, we have also to consider that is ecologically highly plausible that this effect is stronger for the footstep sounds. In fact, in daily life this sound is connoted by a huge variability (i.e. different kind of shoes in different kind of floor produce very different sounds). In other words, it is plausible that the brain would be used to allow a bigger variability respect to this sound in comparison to the hand clapping and, even more, to the tongue click. Finally, when we contrasted the tongue click, hand clapping and footsteps deviants against the water drop events we observed a significant difference between them (all $p_s < 0.05$). Again, this data are in line with our previous conclusion (see experiment 1). There would be two different domains, first, a general motor amenability domain, or motor classification, further specified by a particular motor mapping, or motor specification, inner to the more general motor classification. We speculate that this particular motor specification would operate with mechanisms that are very close to what happens in the tonotopic system during the acoustic stream segregation. This could somehow determine the interaction Sound × Gradient ($F_{6,54}$ = 4.13, $p = 0.029$, $\eta p^2 = 0.31$). Here, the tongue click and hand clapping events produced MMN amplitudes that were similar at frontal and central electrodes ($p > 0.32$), while the footstep sounds showed a more anterior distribution ($p < 0.002$).

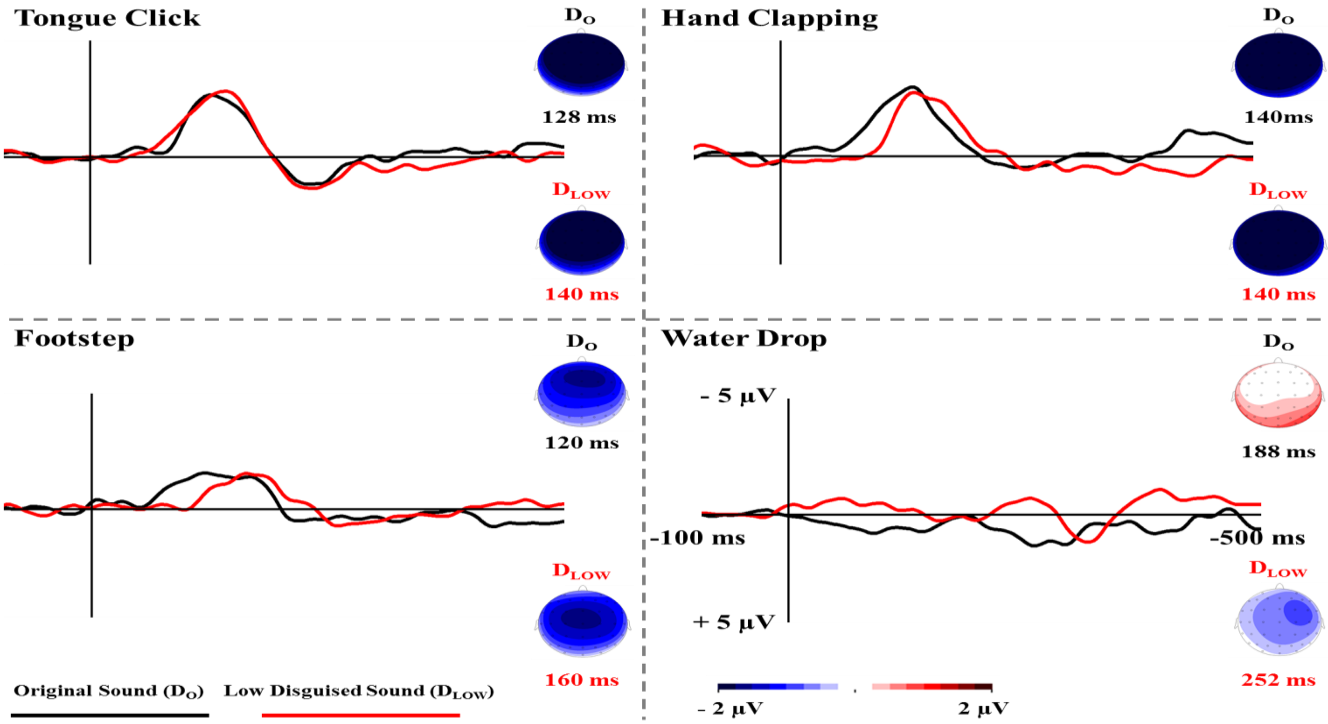


Figure 3.6: Grand averages of nose-referenced deviant-minus-repeated difference signals of 10 subjects at FCz electrode for all sounds that entered in the final analysis.

SOURCE ESTIMATION

The main reason below the source analysis was to better understand the difference between the original sounds and their respective D_{LOW} events. To this end we carried out paired t-tests to directly compare the source estimation of each Original sound with its relative D_{LOW} sound. Noteworthy, this analysis showed that the only one significant difference has been found in relation to the Tongue click sounds. In particular, when we contrasted the original sound against the D_{LOW} we observed two bilateral significant clusters (cluster level both $p_s < 0.03$ FDR corrected; MNI coordinate 52, -10, 18 and -56, -32, 12; see Figure 3.7). Although the poor spatial resolution of source estimation do not allow a precise localization of this source, it is interesting to note that this map refers to the stronger activation elicited by the original tongue click sound when compared with its low disguised. Therefore, could be possible that this stronger activation would include activity from the bilateral mouth representations. Importantly, this data have been previously reported in an action perception MMN study (Hauk *et al.*, 2006).

Figure 3.7: Contrast Original tongue click sound against tongue click D_{LOW} sound. Significant clusters ($p < 0.03$, corrected)

Therefore, since we did not observe any other significant effect when the original sounds were contrasted against their low disguised sounds, could be possible that the D_O and D_{LOW} events were completely segregated only in the most extreme situation (i.e. tongue click, see above VAS scores). As previously mentioned, it is highly plausible that the brain allows less degree of freedom to this sound category. In other words, in comparison to the hand clapping and, even more, to the footstep sounds the brain is used to associate less changing to this particular stimuli (i.e. all the tongue click sounds are more similar to each other than the footsteps are). This fact would further, strength our interpretation, that is, more "distant" two sounds are within the motor amenability dimension more segregated they would be regard the particular motor specification.

3.4 Discussion

When listening to someone at a crowded cocktail party, or trying to follow the second violin line in a symphonic orchestra, we rely on our ears' and brain's extraordinary ability to parse complex acoustic scenes into individual auditory "objects" or "streams" (Griffiths and Warren, 2004). This ability is strictly related to the main role of perception, that is, understanding which events in the outside world caused the sensory observations. However, the everyday auditory environment consists of multiple simultaneously active sources with overlapping temporal and spectral acoustic properties. Our brain, therefore, has to segregate this mixture of the concurrent sound streams from each other and to attribute them to their original sources. Assigning sensory inputs to the sound sources they belong to (that is, building a neural representation of the auditory environment) is called auditory stream segregation (Bregman, 1990). Evidence for the existence of unintentional encoding of auditory rules, by which organizing sounds, has been provided by several passive oddball studies using the Mismatch Negativity brain wave of the event-related potential (ERP) (Carral *et al.*, 2005; Saarinen *et al.*, 1992; van Zuijen *et al.*, 2005; Zachau *et al.*, 2005). These rules are fundamental in everyday environments where the sound sources often generate series of discrete sounds, (e.g. footsteps). A substantial part of the information characterizing the source, such as whether a person is approaching or receding, is not present separately within the individual sounds: it can be extracted only by relating the individual sounds to each other. In the present study we investigated whether similar rules mediate the sound into action translation process. At this regard our results strongly support the view that during action perception the brain is able to make arbitrary association between sounds. In particular, the results from the first experiment seem indicate that the presence of intelligible action sounds increases the likelihood that auditory-motor associations extend to less intelligible ones. Noteworthy, this mechanism seems selective for the abstract motor amenability dimension. In fact, the multideviant blocks have been designed in such a way to reduce as much as possible any acoustic inference in this process. This result seems in line with increasing evidence in the psychoacoustic research field, where has been demonstrated that the acoustic system accommodate each incoming sound in previous acoustic representation. This mechanism of accommodation would mainly be based on the relation between the incoming sound and the previous events (Friston, & Kiebel, 2009; Winkler, 2007). Therefore, the results from the first experiment seem show that the brain 'accommodates' the stimuli on the basis of a general 'motor' heuristic. Furthermore, the second experiment tell us that within the same general heuristic some kind of weighting occurs depending on the 'subjective distances' of the deviants. In particular when the deviants were assessed more similar to each other (i.e. footstep stimuli) we observed a reduction of the MMN mean amplitude in comparison to the conditions in which this events were more separated (tongue click and hand clapping). As previously mentioned, this pattern of results is in line with the stream segregation phenomenon (Bregman 1990). When two sounds are fused together their discrimination became harder and, hence, the MMN smaller. Vice versa, when two sounds are segregated their discrimination is easier and, hence, bigger is the MMN (Müller *et al.*, 2005; Sussman *et al.*, 1999). Therefore, the coupling effect would be modulated by the relative distance of the sounds. Above all that, even if the D_{LOW} belongs to an action representation, there are still questions to be answered: does it tell us anything new about the behavior of the sound source (i.e. body source), or is it entirely predictable for us and thus simply confirms what we already know? Our speculation is that a stream segregation-like mechanism would have an important role in solving this uncertainty. In doing this, the sensorimotor cortices would operate as the acoustic system does in solving similar perceptual problems. At this regard, most of the neurons of the auditory system, where usually the acoustic stream segregation is investigated, are frequency selective. Therefore, has been demonstrated in the primary auditory cortex (Area A1) of awake macaques that the response of a neuron tuned to one specific frequency A will extend to a very similar frequency B (small Δf), when this two tone frequencies are presented in sequential and alternated order (i.e. A B A B). However, as Δf is increased, that neuron responded only to the A tones, other neurons would respond to the B tones, but very few would respond to both (Micheyl, *et al.*, 2003; Fishman, *et al.*, 2001). Therefore, could be possible that the somatotopic organization would work in a very similar way. In particular, while the original sounds activate the coherent sensorimotor representations (for a review Aglioti & Pazzaglia 2010), the D_{LOW} events would be cortically encoded on the basis of their perceptual relation with the original sound. However, could be possible that as the perceptual distance decrease a mutual influence would emerge. At this regard, the source analysis showed a significant difference only for the tongue click sound category where the two deviants were perceptually well-separated. One possible explanation could be that in the condition in which the two deviants were assessed more close to each other (i.e. hand clapping and footstep), the corresponding sources became progressively closer to each other and, hence, no difference emerged. Alternatively, one could speculate that in the condition in which maximum was the distance (i.e. the tongue click) the original sound could elicit a well-defined cortical source. The D_{LOW}, instead, did not show any clear sensorimotor cortical activation because it would elicit a spreader sources with a huge variability both inter trials and between subjects. Instead, in the more intermediate condition (i.e. hand clapping and footstep sound) where the perceptual salience of the D_{LOW} events was higher, the two sounds could influence each other and, therefore, not well-defined the corresponding sources were. However that may be, the specific relation between the two sounds seems play an important role coherently with the analogy with the auditory perceptual analysis.

Chapter IV

The role of massive somatic deafferentation and motor deefferentation of the lower and upper part of the body in audio-motor mapping

4.1 Introduction

Embodied cognition theories postulate that perceiving and understanding the body states of other individuals are underpinned by the neural structures activated during first-hand experience of the same states. This suggests that one's own sensorimotor system may be used to identify the actions and sensations of others. Virtual and real brain lesion studies show that visual processing of body action and body form relies upon neural activity in the ventral premotor and the extrastriate body areas, respectively. In particular, transient inactivation (Urgesi *et al.*, 2004) or permanent lesion (Moro *et al.*, 2008) of the extrastriate body area impairs the analysis of purely visual properties (e.g. the form) of non-facial body parts. In a similar vein, transient (Avenanti *et al.* 2007; Urgesi *et al.*, 2007a,b; Candidi *et al.*, 2008) or permanent (Moro *et al.*, 2008; Pazzaglia *et al.*, 2008; Serino *et al.*, 2010; Aglioti & Pazzaglia 2011) inactivation of the premotor and parietal cortices impairs the sensorimotor mapping of seen bodily actions. In the next study we explored whether the perception of action-related sounds may also be altered in the absence of damage to the above cortical regions by testing, with the same paradigm tested in the previous chapters, healthy controls and spinal cord injury (SCI) patients whose brain was unable to receive somatic information from and send motor commands to the body parts below the lesion level. At this regard, has been recently demonstrated (Pernigo *et al.*, 2012) that SCI patients have a specific, cross-modal deficit in the visual recognition of the disconnected lower body parts. This deficit affected both body action and body form perception, hinting at a pervasive influence of ongoing body signal on the brain network dedicated to visual body processing. Therefore, testing this patients could represent a further evidence of the bodily information in auditory motor mapping. Furthermore, it is interesting to test our paradigm in order to see what happens to the D_{LOW} events. In particular since we assume that this class of stimuli are, in comparison to their original sounds, mapped in a different way on the motor system (i.e. the D_O and D_{LOW} would be segregated regard the particular motor specification), we could get more precise indication about the way of encoding of this stimuli.

4.2 Experiment 3: Sound into action translation in spinal cord injured (SCI) patients: a mismatch negativity (MMN) study

Materials and Methods

PARTICIPANTS

Control group: Nine participants (9 males; mean 34.3 range 26 – 50 years, all right handed), with no history of neurological psychiatric or hearing impairment were included in the control group.

Paraplegic group: Nine participants (8 males; mean 36 range 19 – 50 years), were included in the paraplegic group. All of these participants had undergone a traumatic lesion below the skeletal level of the third thoracic vertebra (T3), in the absence of head trauma. The neurological levels of lesions were assessed by means of the American Spinal injury Association Scale (AIS) (Ditunno *et al.*, 1994). Completeness of the neurological lesion was assessed according to the International Standards for Neurological Classification of Spinal Cord Injury, and an absence of sensory and motor function in the lowest sacral segments (S4-S5) was found (Waters *et al.*, 1991); that is all of the patients in the sample scored “A” in the AIS. The average onset of traumatic event was 4.7 years (range 1 – 16 years) before testing.

Tetraplegic patients: Nine participants (9 males; mean 32.1 range 22 – 41 years), were included in the tetraplegic group. All of these participants had undergone a traumatic lesion above the skeletal level of the sixth cervical vertebra (C6), in the absence of head trauma. The neurological levels of lesions were assessed by means of the American Spinal injury Association Scale (AIS). Completeness of the neurological lesion was assessed according to the International Standards for Neurological Classification of Spinal Cord Injury, and poor sensory and motor functions in the highest and lowest segments was found (Waters *et al.*, 1991); that is, two patients in the sample were scored “D”, one was scored “C”, while the remaining were scored “A” in the AIS. The average onset of traumatic event was 8.5 years (range 1 – 22 years) before testing.

STIMULI AND PROCEDURE

The experimental procedures, the subjective rating, the electrophysiological recording methods and the preprocessing stages were identical to those exposed in Experiment 2 (see previous chapter). However, we decreased the length of the multideviant blocks. Each of the deviants with the extreme peak frequency (i.e. D_O and D_{LOW}) were presented 110 times, whereas the $S_{1:4}$ were presented 140 times. The SOA was constant at 400 ms. On average approximately 3 (range 2 – 6) components were removed from each subject in all groups. Furthermore, approximately 5% (range 3 – 6%) of the trials were rejected because they violated the artifact criteria in all groups.

STATISTICAL ANALYSIS

Potential effects of deviant sound processing were assessed for MMN amplitudes at channel FCz, referenced to the electrode placed on the nose, by a mixed ANOVA design with the factors: Sound (i.e., tongue click, hand clapping, footstep, and water drop) and Deviant (i.e., D_O and D_{LOW}) as within factor and Group as between factor (i.e. control, paraplegic, tetraplegic). The subjective ratings of auditory familiarity and perceived motion in the human and non-human action sounds were compared using three separate mixed ANOVAs. The factors were Sound (tongue click, hand clapping, footstep and water drop) and Deviant (D_O, D_{LOW} and D_{HIGH}1:4) as within factor and group as between factor.

RESULTS

VISUAL ANALOGIC SCALE

Repeated measures ANOVAs has been carried out separately per each rating dimension ((1) “How do you think this sound is reproducible by human body performing movement or action?” (2) “How frequently do you produce similar sounds performing movement or action?” (3) “How do you think this sound is reproducible by water? Or, How is it water-related to you?). This analysis revealed, per each rating dimension, a significant main effect of Sound (all $F_{3,72} > 5.43$, $p < 0.003$, $\eta p^2 > 0.18$) and Deviant (all

$F_{5,120} > 10.97$, $p < 0.001$, $\eta p^2 > 0.30$). However, these main effects are not very informative because they refer to the average of very different sounds and deviant. Furthermore, the Sound × Deviant interaction (all $F_{15,360} > 5.70$, $p < 0.001$, $\eta p^2 > 0.19$) was observed in all three rating dimensions (Figure 4.1 and 4.2). Newman–Keuls post-hoc test revealed that, for their respective rating dimension, the original sounds were significantly better perceived as sounds feasible as human or non-human actions than their respective D_{LOW} and $S_{1:4}$ sounds (all $p_s < 0.001$). This was not true just for the original footstep sound that resulted not significantly different in comparison to its D_{LOW} event in the visual analogue scale assessing how frequently the participants produce similar sounds performing actions ($p > 0.14$). However, this result is presumably affected by the lower scores of the patients to the footstep sound. Furthermore, as observed in the previous experiment (see above), the D_{LOW} sounds were significantly better perceived as feasible actions than their respective $S_{1:4}$ (all $p_s < 0.007$) only for the hand clapping and the footsteps sound. The tongue click and water drop D_{LOW} sounds, for their respective rating dimension, showed the same low salience observed in the previous experiment (comparison with their respective $S_{1:4}$ all $p_s > 0.40$). Overall, the results are similar to those previously presented and, hence, coherent with the absolute degree of filtering applied to this class of stimuli.

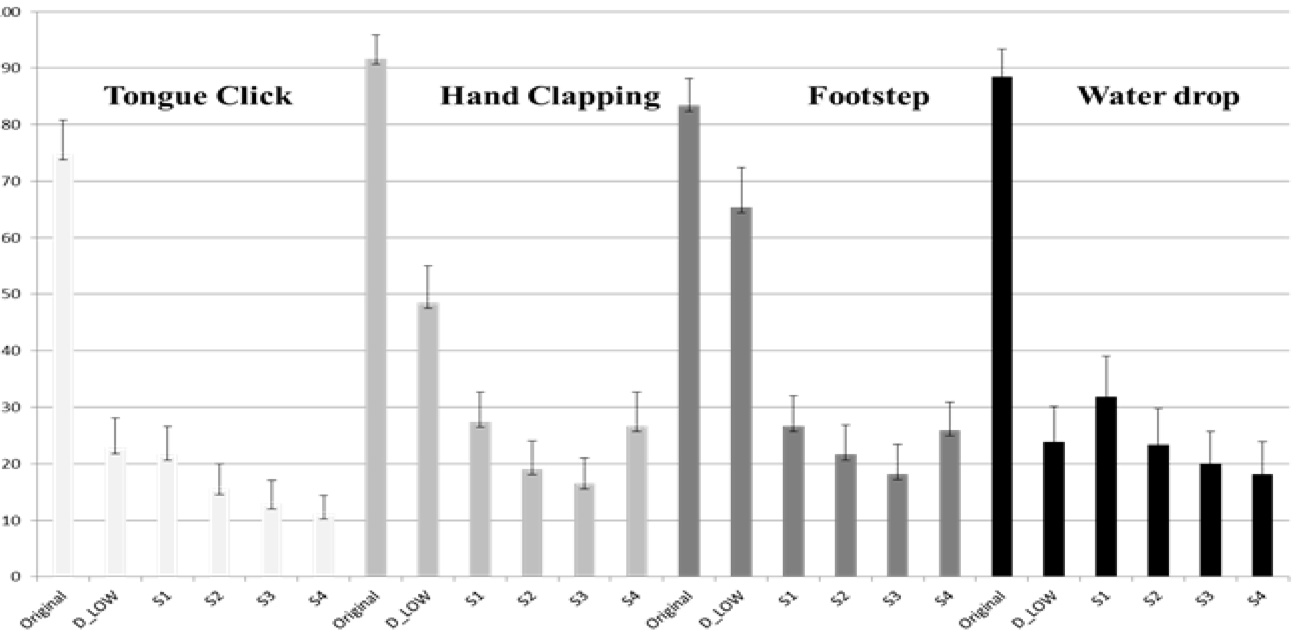


Figure 4.1: Mean + SEM ratings (scale 0-100) on their respective rating dimension. The first eighteen scores refers to the human actions related sounds for their respective rating dimension (VAS 1, see above). The last seven bars refer to the water drop sound (VAS 3)

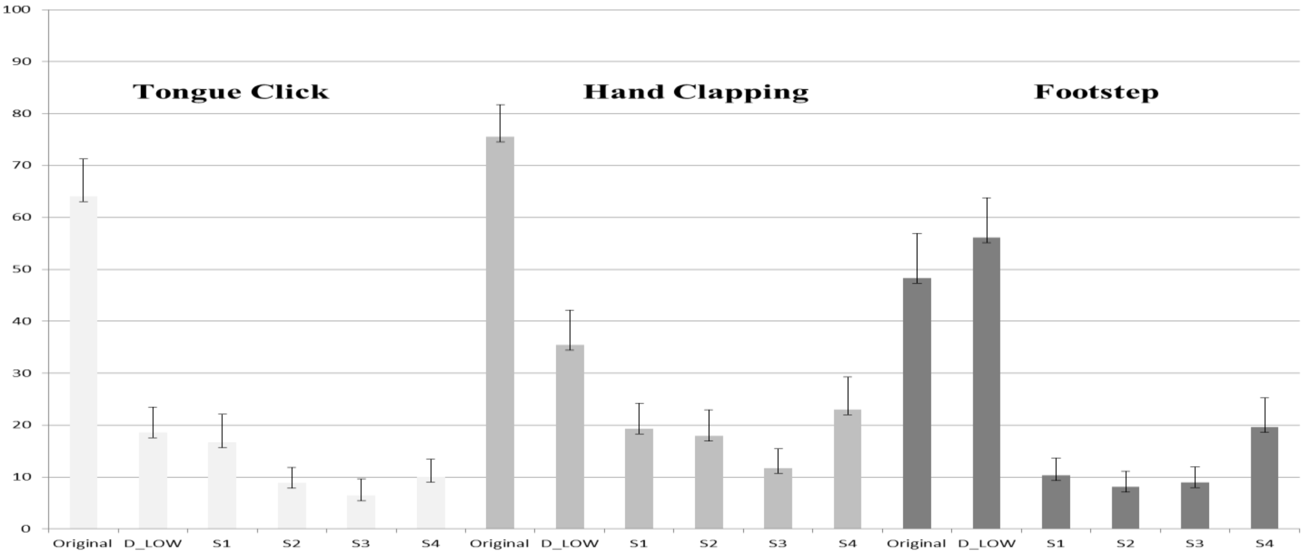


Figure 4.2: Mean + SEM ratings (scale 0-100) on their respective rating dimension. The scores refers to the human actions related sounds for their respective rating dimension (VAS 2, see above).

However, since here we are interested to investigate the role of sensori-motor disconnection in action perception, we are particularly interested to look at the differences between groups on these scales. At this regard, we observed the interaction of factors Sound and Group just for the second visual analogue scale ($F_{6,72} = 2,24$, $p = 0.048$, $\eta p^2 = 0.16$). Although it is highly plausible that we did not observe group differences on the general visual analogue scales (i.e. the VAS 1 and 3), the post-doc test on this interaction revealed just a significant difference whithin the paraplegic group in which the hand clapping sound category was assessed higher than the footstep sounds ($p < 0.026$). Therefore, to elucidate this data, we carried out, on this rating dimension, an ANOVA with the original human action sound scores (i.e. tongue click, hand clapping and footstep D_O) as within variable and group as between factor. In this case the tongue click would be the control sound. In fact, we did not expect differences between groups on this particular sound. Contrary, the hand clapping and footstep sounds should show differences. In particular since the paraplegia refers to a strong sensori-motor impairments for the lower limb and the tetraplegia for the lower and upper limb, we would expect differences on this specific sounds. The results of this ANOVA showed a significant main effect of Group ($F_{2,24} = 4.40$, $p = 0.023$, $\eta p^2 = 0.27$). Newman–Keuls post-hoc test revealed that, both controls and tetraplegics gave higher scores than the paraplegic group (both $p_s < 0.035$). However, the control and tetraplegic groups were not different between them ($p = 0.59$). This result seems contradictory, since we would expect also a difference between tetraplegic and control participants. Furthermore, the interaction Sound x Group was observed ($F_{4,48} = 4,12$, $p < 0.006$, $\eta p^2 = 0.26$). This interaction showed that the three groups were not different in relation to the tongue click sound (all $p_s > 0.51$) and, crucially, to the hand clapping sound (all $p_s > 0.68$). Finally, the paraplegic patients gave lower scores to the footstep sound in comparison to the healthy subject ($p = 0.001$) and, crucially, they showed a strong tendency to differ also in comparison to the tetraplegic patients ($p = 0.054$) (Figure 4.3). These results seem indicate that when directly asked about the probability to produce similar sounds performing action in daily life, the tetraplegics exhibited an overall tendency to overestimate they ability to move. Although this results are not sufficient to draw precise and safe conclusions, could be possible that the massive sensorimotor disconnection induced in this patients also a more generalized deficits of awareness or motor monitoring. Although, this kind of deficits are usually reported in relation to hemiplegia after brain injures (i.e. anosognosia for hemiplegia), could be possible that a similar phenomenon is present also in this kind of sensori-motor disorder. At this regard, it is interesting to note that Berti and colleagues (2005) proposed a modular mechanism under the awareness of motor impairments. In particular, in their famous brain lesion analysis study this authors (Berti *et al.*, 2005) showed the highly plausibility that damages to the sensori-motor areas (Brodmann's premotor areas 6 and 44, motor area 4 and the somatosensory cortex) would impairs the motor-monitoring process. We speculate that the tetraplegic patients, who show a massive de-afferentetion and de-efferentation of both the upper and lower limbs, became less accurate in assessing their ability to move. This could be explained in two not mutually exclusive ways, first it is possible that occurred a general re-organization of this areas following this massive disconnection. Alternatively, could be possible that, even without any plastic changes in sensorimotor cortices, this ability became less accurate because of a massive pauperization of the sensorimotor information. Therefore, although they do not show any of the higher denying mechanisms usually associated with anosognosia for hemiplegia (i.e. confabulation), could still be possible that the massive de-afferentation and de-efferentation could somehow impairs their ability to accurately assess their motor competences. However, this data are just an indication. More research should be carried out on this side to better understand the plausibility of this phenomenon.

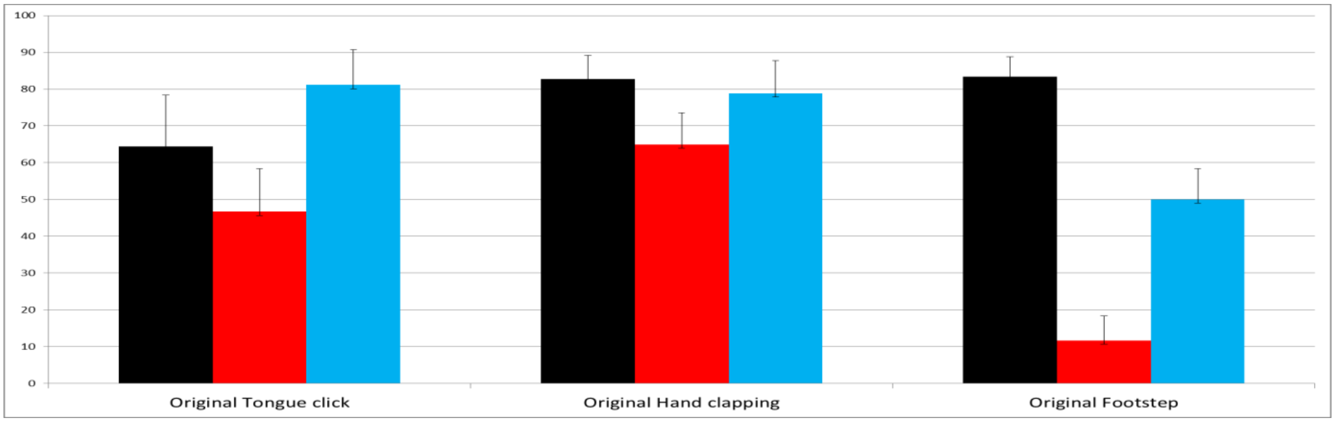


Figure 4.3: Mean + SEM ratings (scale 0-100) for the original human action related sounds on the VAS 2 (see above). Black bars refer to the control group, Red bar to the paraplegic participants, while, the blue ones to the tetraplegic group

MISMATCH NEGATIVITY

The ANOVA on the MMN mean amplitudes revealed a main effect of Sound ($F_{3,72} = 61.17$, $p < 0.001$, $\eta p^2 = 0.71$). Newman-Keuls post-hoc test for multiple comparisons revealed that the tongue click and hand clapping sound categories (D_O and D_{LOW} collapsed and collapsed between groups) were significantly different in comparison to all the other sounds (all $p_s < 0.001$). Furthermore, the tongue click sound category elicited bigger MMN in comparison to the hand clapping sound category ($p < 0.001$). Finally, the footstep sound category elicited MMNs that, even if significantly smaller than the tongue click and hand clapping sound, showed a strong tendency to differ in comparison to the water drop sound category ($p = 0.057$). Overall this mean effect showed a modulation of the MMN depending on sound category that seems to be roughly coherent with the sensori-motor impairments of the spinal cord injury patients. In other words, since the tongue click would not be affected by the lesions it elicits a bigger MMN response in comparison to both hand clapping and footstep sounds. At this regard the Sound X Group and Sound X Deviant X Group interactions have been observed (both $F_{6,72} > 5.78$, $p < 0.001$, $\eta p^2 = 0.32$). Newman-Keuls post-hoc test for multiple comparisons revealed that the tongue click original sound elicited MMNs that were similar between control and paraplegic ($p = 0.91$) participants, but bigger in the tetraplegic group (both $p_s < 0.04$). This data is particularly interesting because seems indicate that for this sound the tetraplegic patients are more sensible in comparison to the other groups. We speculate, that this data could suggest a cortical reorganization for these patients who have massive sensori-motor disconnections for both lower and upper limb (lesion above C6 vertebra). Noteworthy, the tongue click D_{LOW} sound elicited in the paraplegic group a bigger MMN in comparison to both control and tetraplegic groups (both $p_s < 0.032$), that were not statistically different between them ($p = 0.98$) . This data is particularly interesting because it would suggest that the specific audio-motor mapping for this special class of sounds is sensible to the level of sensori-motor disconnection. If we consider the previous hypothesis stating that the brain assigns the particular motor specification to this sound category depending on the relationship with its original we could draw some conclusions. Because, we speculated that the original and D_{LOW} events would be segregated within the particular audio-motor mapping, it is highly plausible that the MMN elicited by the original tongue click in the tetraplegic group and that one elicited by the tongue click D_{LOW} in paraplegics could share a common spreader encoding due to the skeletal lesion. Further, insights to this fact derive from the comparison of the hand clapping sounds. Here, the MMN elicited by the original hand clapping sound was different between control and tetraplegic, and between paraplegic and tetraplegic (both $p_s < 0.008$), while controls and paraplegics were not different ($p > 0.87$). These results seem indicate that the skeletal lesion level and its consequence on the sensori-motor disconnection, could somehow affects the amplitude of this component. Crucially, when we compared the hand clapping D_{LOW} event we observed a drop of the MMN mean amplitude for both paraplegic and tetraplegic in comparison to the control group (both $p_s < 0.009$), that were not significantly different between them ($p = 0.99$). This results is particularly interesting because would suggest that the particular audio motor mapping (or specification) of this special class of stimuli is not “completely” related to the original one but, instead, its cortical implementation would be well separated (i.e. segregated). Furthermore, it could be plausible that for this sound the implementation would interest some action representation partially distorted or silent in the paraplegic group (e.g. thoracic or lower limb). Furthermore, the D_O and D_{LOW} footstep sounds for both paraplegic and tetraplegic showed smaller MMNs in comparison to the MMN elicited by the same sounds in the control group (all $p_s < 0.04$). Finally, no significant differences have been found between groups in relation to the water drop events (all $p_s > 0.89$) (from Figure 4.4 to Figure 4.7).

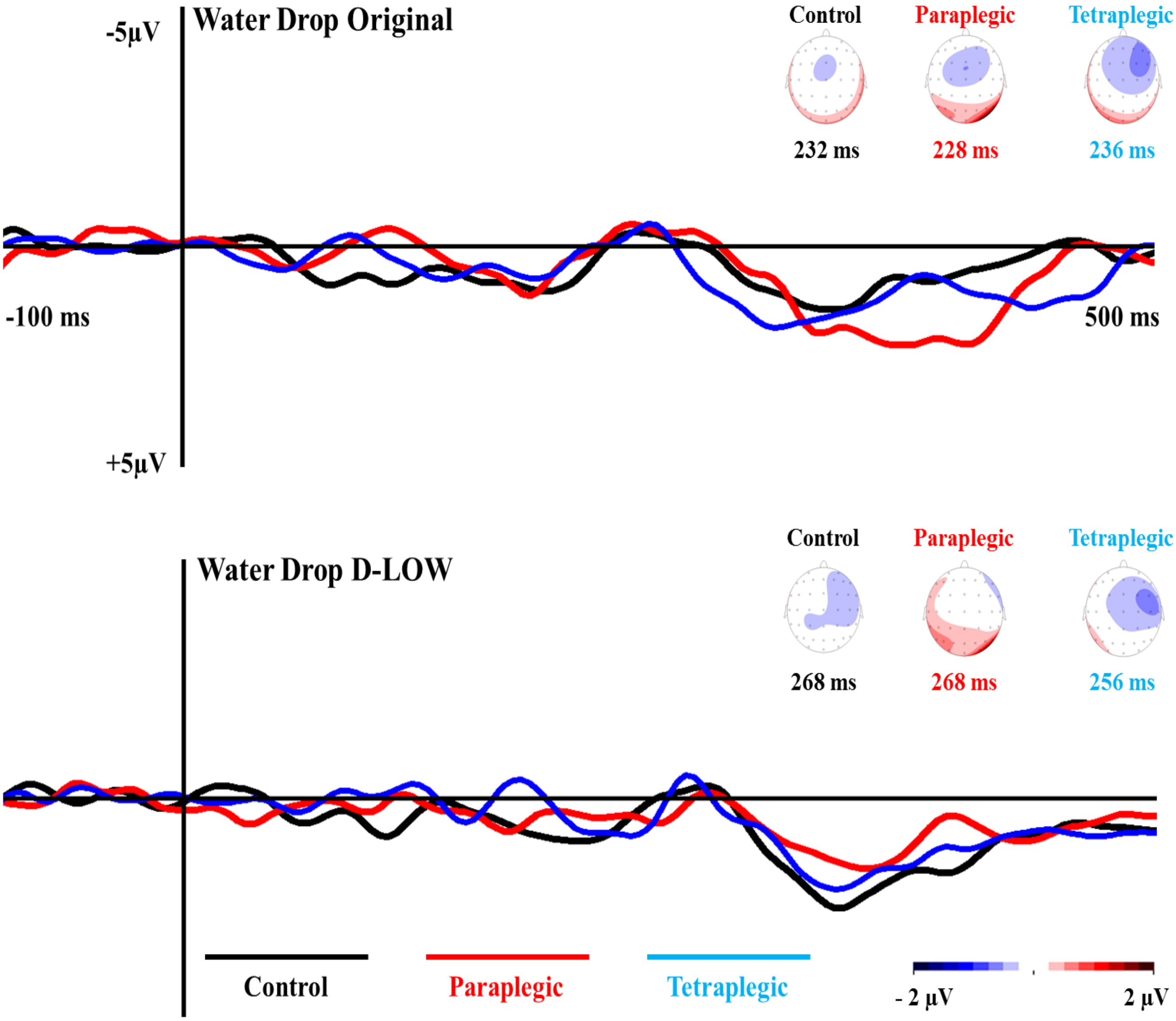


Figure 4.4: Grand averages of nose-referenced deviant-minus-repeated difference signals at FCz electrode for the water drop sounds.

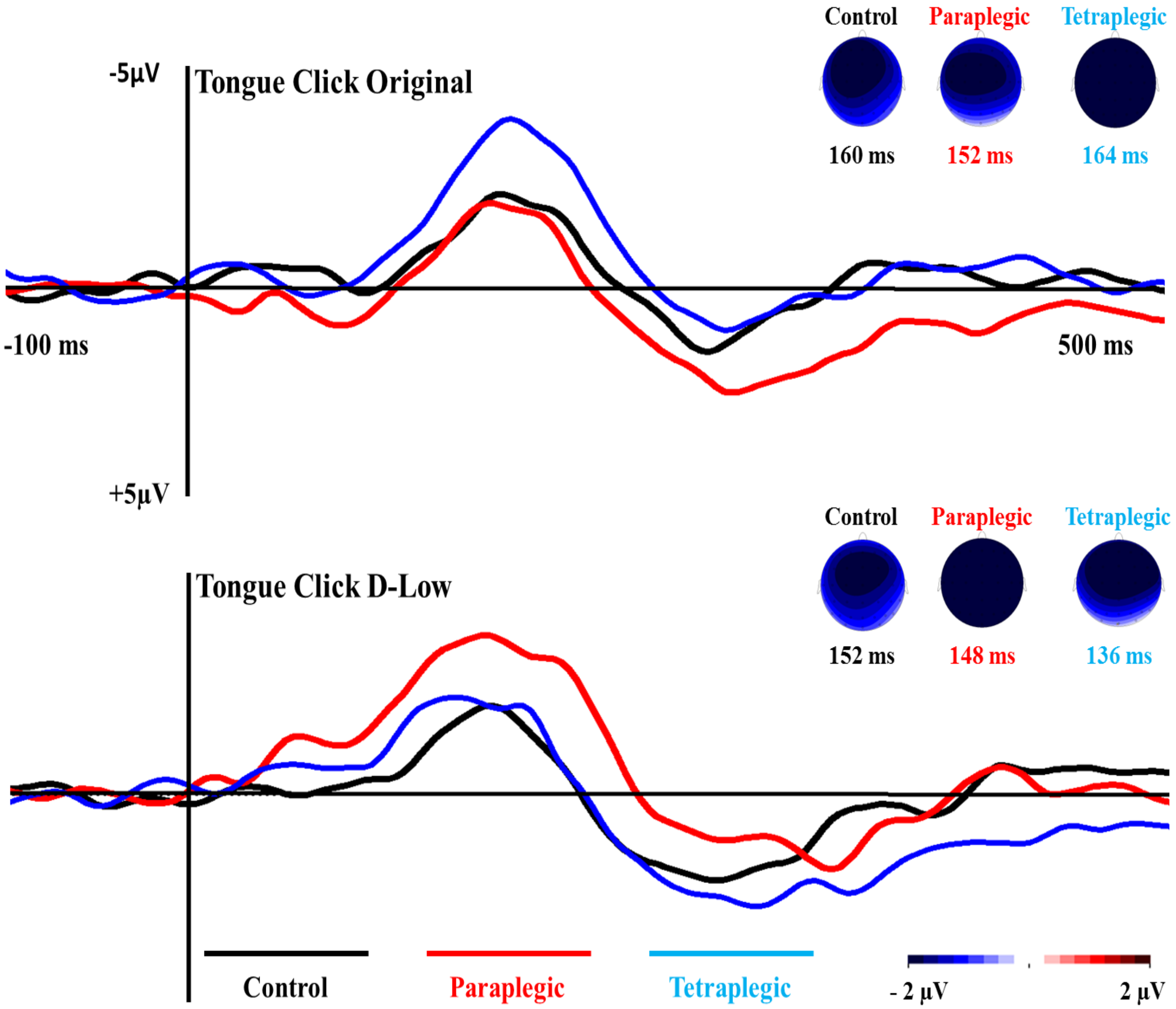


Figure 4.5: Grand averages of nose-referenced deviant-minus-repeated difference signals at FCz electrode for the tongue click sounds.

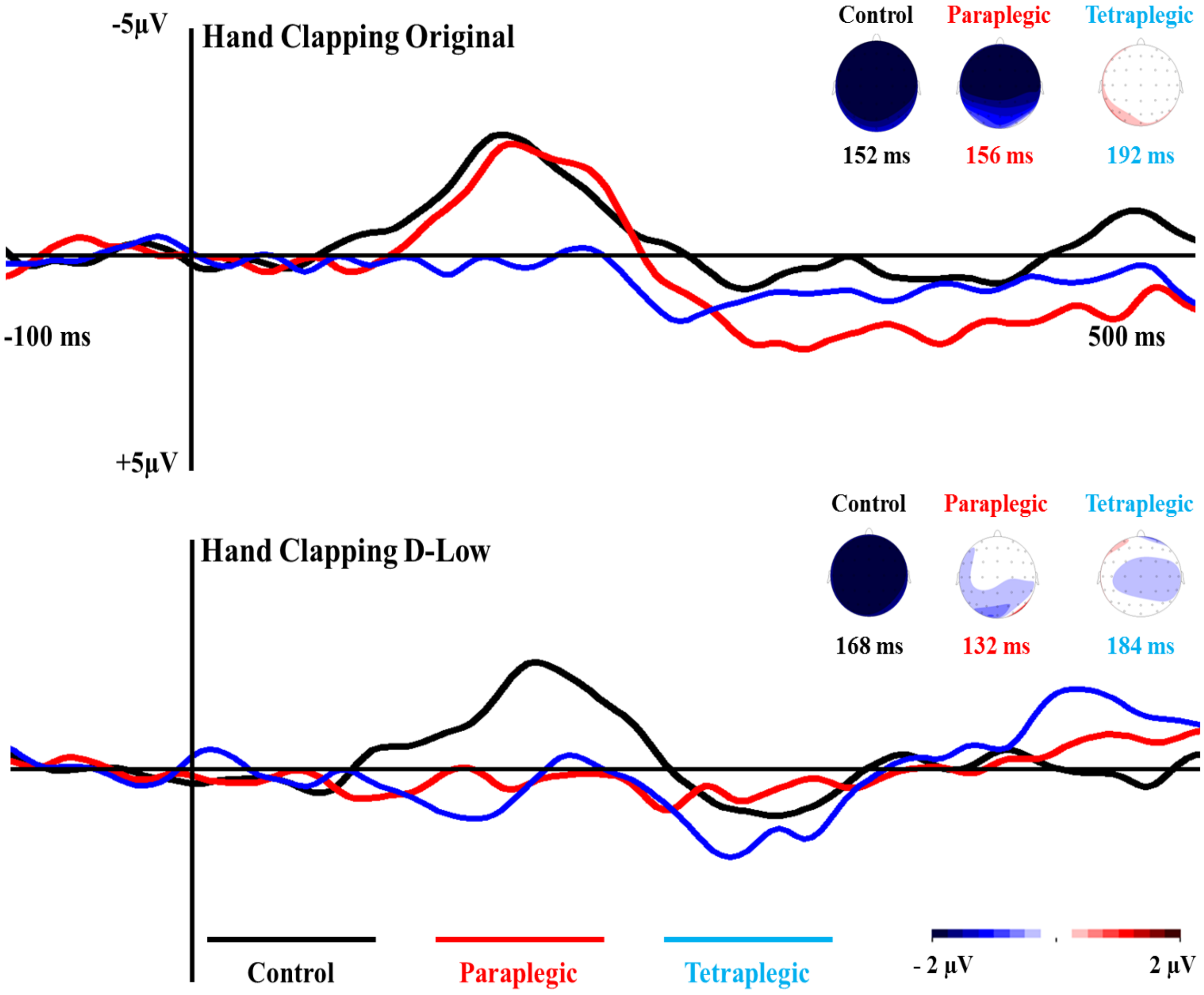


Figure 4.6: Grand averages of nose-referenced deviant-minus-repeated difference signals at FCz electrode for the hand clapping sounds.

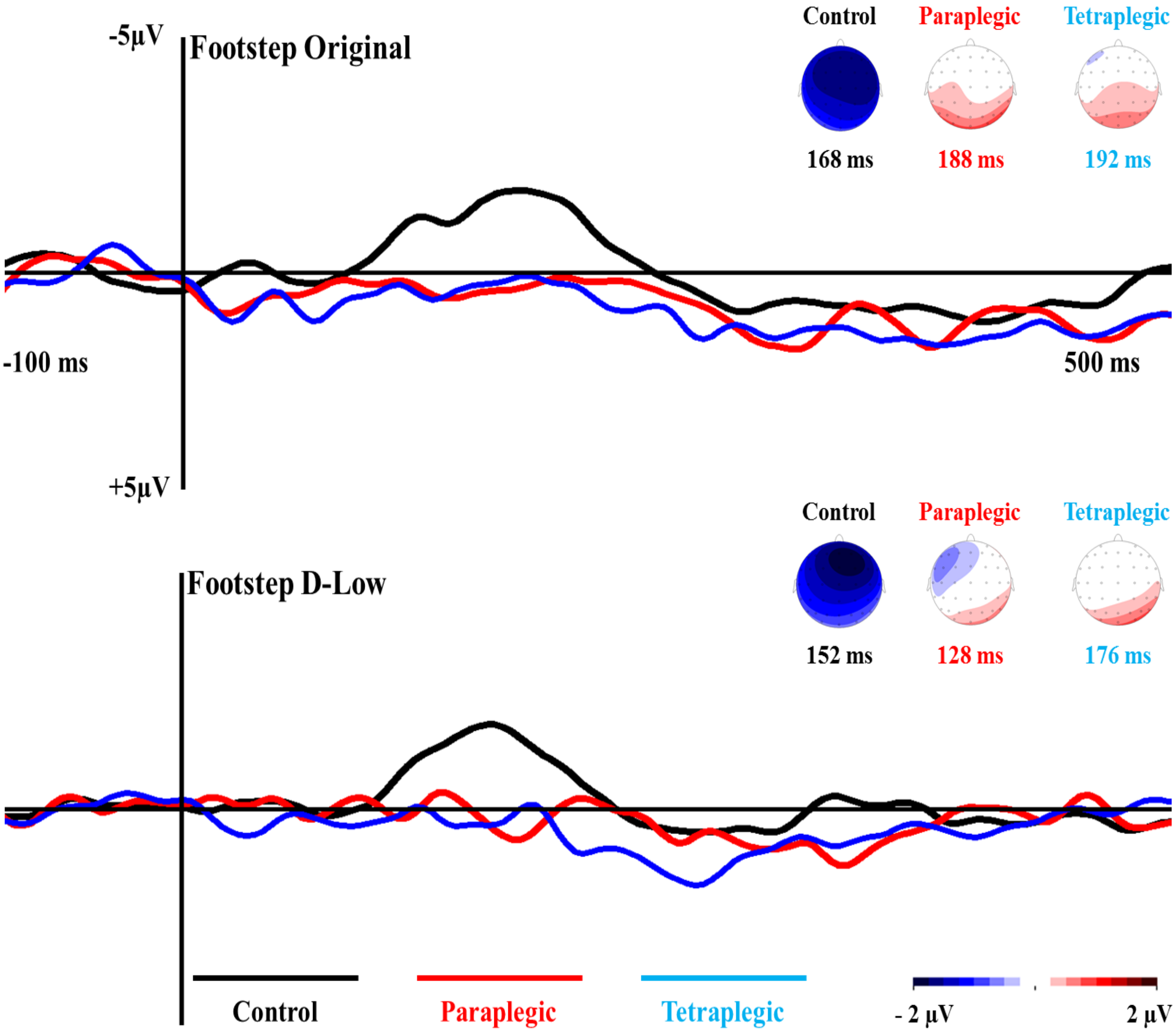


Figure 4.7: Grand averages of nose-referenced deviant-minus-repeated difference signals at FCz electrode for the footstep sounds.

Discussion

The acoustic perception of human action related sounds relies upon neural activity in cortical areas that deal with the receipt of somatic afferences from the body and the issuing of motor commands to it. Indeed, functional neuroanatomy studies indicate that mere observation of body actions activates fronto-parietal somatosensory and motor cortices (Buccino *et al.*, 2001; Avikainen *et al.*, 2002; Rossi *et al.*, 2002; Raos *et al.*, 2004; Costantini *et al.*, 2005; Iacoboni *et al.*, 2005; Avenanti *et al.*, 2007; Christensen *et al.*, 2007; Sakamoto *et al.*, 2009). Thus, the somatic and motor representations of observed body actions may feed back into perceptual areas (e.g. visual) and affect perceptual processing by providing the visual system with a fine-grained description of the static and moving body (Wilson & Knoblich, 2005). In this study we got more information of these mechanisms. In particular, two aspects are important. First we observed that the original human action related sounds were strongly affected by the level of sensori-motor impairments. In particular, it is interesting to note the high level of selectivity of this results. At this regard, we observed that the paraplegic group showed a significant decrease of MMN only in relation to the footstep original sounds, while the tetraplegic group showed a similar pattern also for the hand clapping sound. It is highly plausible that this results are directly determined by the specific sensori-motor impairments of this patients. Noteworthy, the tetraplegic group showed a significant increasing of the MMN response in relation to the tongue click. At this regards, one speculation could, carefully, drawn. It is possible that the enhanced MMN response to this particular sound could be due to the involvement of surrounding “silent” areas (e.g. hand representation). This, explanation could also explain why this sound category did not show a similar MMN also for the D_{LOW} event. In fact, we assume (as the previous experiment would suggest, see above) that the D_O and D_{LOW} events would be determined (segregated or fused) on the basis of a process of mutual “negotiation” in which the main variable is the perceptual distance between this sounds (see Experiment 2 above). Therefore, since the original tongue click sound would activate a spreader areas in the tetraplegic group than in the control and paraplegic groups (where the surrounding mouth and hand areas are well connected with their respective body parts) the D_{LOW} sound, perceptually distant, would not be encoded with the same motor codification and, hence, it would activate a different and smaller area. The specular results in the paraplegic group (i.e. a bigger MMN for the tongue click D_{LOW} events in comparison to its original sound) would, further, support this hypothesis. Here, contrary to the tetraplegic results, the original tongue click sound would be well segregated in its physiological action representation while, the tongue clicks D_{LOW} events would be mapped on spreader motor activation. Therefore, could be possible that for some reason this particular event triggers a spreader, or stronger, cortical responses in this group.

A second interesting results, that further support our idea of segregated motor response for the D_O and D_{LOW} events, come from the hand clapping MMNs in the paraplegic group. Here we observed a drop of the MMN for the hand clapping D_{LOW} sound in comparison to its original. Could be possible that this sound would be segregated (on the basis of perceptual dissimilarity) on some areas particularly affected by the lesion level of these patients (e.g. lower limb). Therefore, source analysis also on this experiment will be carry out with the intent to further support the above mentioned hypothesis.

General Discussion

The ability to derive the intentions of others from the sound produced by their actions is quintessential to effective social behaviour. Many neuroscientists believe that this ability depends on the brain’s mirror-neuron system, which provides a direct link between action and perception. Precisely how intentions can be inferred through action-perception, however, has provoked much debate. One challenge in inferring the cause of a perceived action, is the fact that the problem is ill-posed, because identical movements can be made to perform different actions with different goals. Here, we show how, in the auditory modality, identification of most likely cause of a human action-related sound is highly subject to inferences. Using multi-channel, event-related potentials (ERPs), we determined the temporal dynamics of the ability to decipher action sounds by recording the mismatch negativity (MMN) generated in response to multi-deviant stimuli consisting of 3 different human action-related sounds (click of the tongue, hand clapping, and footsteps) and a non-human action-related sound (water drop). Subjects listened to the original sound-stimulus and to sounds obtained by altering 1 (low degree of disguise) or more complex (high degree of disguise) acoustic parameters of the original sound. Overall, the results indicate that the presence of intelligible action sounds increases the likelihood that auditory-motor associations extend to less intelligible sounds. This automatic mechanism may serve the early perception of action-related sounds in ‘noisy’ environments. Furthermore, the same experiment carried out with spinal cord injured (SCI) patients suggest that massive somatosensory and motor disconnection between the body and the brain may induce functional cortical and subcortical changes, particularly in the regions involved in somatosensory and motor processing.

References

- Aglioti SM, Pazzaglia M (2010) Representing actions through their sound. Exp Brain Res
- Aglioti, S.M. & Pazzaglia, M. (2011) Sounds and scents in (social) action. Trends Cogn. Sci., 15, 47–55
- Alain C, Arnott SR, Picton TW (2001) Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. J Exp Psychol; 27:1072–89
- Alain C, Schjuler BM, McDonald KL. (2002) Neural activity associated with distinguishing concurrent auditory objects. J Acoust Soc Am; 111:990–5
- Alain C, Woods DL. Signal clustering modulates auditory cortical activity in humans (1994) Percept Psychophys; 56:501–16
- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. Exp Brain Res 166:559–571
- Amenedo E, Escera C (2000) The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. Eur J Neurosci; 12:2570–4
- Anstis, S., Saida, S., (1985) Adaptation to auditory streaming of frequency- modulated tones. J. Exp. Psychol. Hum. Percept. Perf. 11, 257-271
- Avenanti A, Bolognini N, Maravita A, Aglioti SM (2007) Somatic and motor components of action simulation. Curr Biol 17:2129–2135
- Avikainen S., Forss N., Hari R., (2002) Modulated activation of the human SI and SII cortices during observation of hand actions. Neuroimage, 15, 640-646
- Aziz-Zadeh L, Iacoboni M, Zaidel E, Wilson S, Mazziotta J (2004) Left hemisphere motor facilitation in response to manual action sounds. Eur J Neurosci 19:2609–2612
- Aziz-Zadeh L, Wilson SM, Rizzolatti G, Iacoboni M (2006) Congruent embodied representations for visually presented actions and linguistic phrases describing actions. Curr Biol 16:1818–1823
- Barsalou LW (2008) Grounded cognition. Annu Rev Psychol 59:617–645

- Berti A, Bottini G, Gandola M, Pia L, Smania N, Stracciari A, *et al.* (2005) Shared cortical anatomy for motor awareness and motor control. Science, 309, 488–91

- Bregman, A.S. (1990) Auditory Scene Analysis: the Perceptual Organization of Sound. MIT Press, Cambridge, MA

- Bregman, A.S., Campbell, J. (1971) Primary auditory stream segregation and perception of order in rapid sequences of tones. J. Exp. Psychol. 89, 244-249

- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ (2001) Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur J Neurosci 13:400–404

- Buccino G, Vogt S, Ritzl A, Fink GR, Zilles K, Freund HJ, Rizzolatti G (2004) Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. Neuron 42:323–334

- Caetano G, Jousma¨ki V, Hari R (2007) Actor's and observer's primary motor cortices stabilize similarly after seen or heard motor actions. Proc Natl Acad Sci USA 104:9058–9062

- Candidi M, Vicario CM, Abreu AM, Aglioti SM (2010) Competing mechanisms for mapping action-related categorical knowledge and observed actions. Cereb Cortex [Epub ahead of print]

- Candidi, M., Urgesi, C., Ionta, S. & Aglioti, S.M. (2008) Virtual lesion of ventral premotor cortex impairs visual perception of biomechanically possible but not impossible actions. Soc. Neurosci., 3, 388–400

- Carral V., Huotilainen M., Ruusuvirta T., Fellman V., Näätänen R., Escera C. (2005) A kind of auditory 'primitive intelligence' already present at birth. Eur Jour Neur, 21, 3201-3204

- Casile A, Giese MA (2006) Nonvisual motor training influences biological motion perception. Curr Biol 16:69–74

- Christensen MS., Lundbye-Jensen J., Geertsen SS., Petersen TH., Paulson OB., Nielsen JB. (2007) Premotor cortex modulates somatosensory cortex during voluntary movements without proprioceptive feedback. Nat. Neurosci., 10, 417-419

- Cohen J (1973) Eta-squared and partial eta-squared in fixed factor ANOVA designs. Educational and Psychological Measurement, Vol. 33(1), 107-112

- Costantini M., Galati G., Ferretti A., Caulo M., Tartaro A., Romani GL., Aglioti SM. (2005) Neural systems underlying observation of humanly impossible movements: an fMRI study. Cereb. Cortex, 15, 1761-1767
- De Lucia M, Camen C, Clarke S, Murray MM (2009) The role of actions in auditoryobject discrimination. Neuroimage 48:475–485 (Epub 2009 Jun 24)
- Decety J, Gre`zes J, Costes N, Perani D, Jeannerod M, Procyk E, Grassi F, Fazio F (1997) Brain activity during observation of actions. Influence of action content and subject's strategy. Brain 120:1763–1777
- Delorme A & Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. Journal of Neuroscience Methods 134:9-21
- di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992) Understanding motor events: a neurophysiological study. Exp Brain Res 91:176–180
- Ditunno, J.F., Young, W., Donovan, W.H. & Creasey, G. (1994) The International Standards Booklet for Neurological and Functional Classification of Spinal Cord Injury. Paraplegia, 32, 70–80
- Fadiga L, Craighero L, Buccino G, Rizzolatti G (2002) Speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur J Neurosci 15:399–402
- Fadiga L, Fogassi L, Pavesi G, Rizzolatti G (1995) Motor facilitation during action observation: a magnetic stimulation study. J Neurophysiol 73:2608–2611
- Fay, R.R. (1998) Auditory stream segregation in gold¢sh (Carassius auratus). Hear. Res. 120, 69-76
- Fishman, Y.I. *et al.* (2001) Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear.Res. 151, 167–187
- Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G et al (2005) Parietal lobe: from action organization to intention understanding. Science 308:662–667
- Friston, K. & Kiebel, S. (2009) Cortical circuits for perceptual inference. Neural Netw. 22, 1093–1104
- Galati G, Committeri G, Spitoni G, Aprile T, Pitzalis S et al (2008) A selective representation of the meaning of actions in the auditory mirror system. Neuroimage 40:1274–1286
- Gallese V, Fadiga L, Fogassi L, Rizzolatti G (1996) Action recognition in the premotor cortex. Brain 119:593–609

- Gangitano M, Mottaghy FM, Pascual-Leone A (2001) Phase-specific modulation of cortical motor output during movement observation. Neuroreport 12:1489–1492
- Gazzola V, Aziz-Zadeh L, Keysers C (2006) Empathy and the somatotopic auditory mirror system in humans. Curr Biol 16:1824–1829
- Gazzola V, Keysers C (2009) The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. Cereb Cortex 19:1239–1255
- Griffiths TD., Warren JD. (2004): What is an auditory object? *Nature Reviews Neuroscience*, 5, 887-892
- Haslinger B, Erhard P, Altenmu¨ller E, Schroeder U, Boecker H, Ceballos-Baumann AO (2005) Transmodal sensorimotor networks during action observation in professional pianists. J Cogn Neurosci 17:282–293
- Hauk O, Johnsrude I, Pulvermüller F (2004) Somatotopic representation of action words in human motor and premotor cortex. Neuron 41:301–307
- Hauk O, Shtyrov Y, Pulvermuller F (2006) The sound of actions as reflected by mismatch negativity: rapid activation of cortical sensory-motor networks by sounds associated with finger and tongue movements. Eur J Neurosci 23:811–821
- Horváth J, Czigler I, Sussman E, Winkler I (2001) Simultaneously active preattentive representations of local and global rules for sound sequences in the human brain. Cogn Brain Res;12:131–44
- Huotilainen M, Ilmoniemi RJ, Lavikainen J, Tiitinen H, Alho K, Sinkkonen J, *et al.* (1993) Interaction between representations of different features of auditory sensory memory. NeuroReport; 4:1279–81

- Huotilainen M, Ilmoniemi RJ, Lavikainen J, Tiitinen H, Alho K, Sinkkonen J, *et al.* (1993) Interaction between representations of different features of auditory sensory memory. NeuroReport; 4:1279–81

- Iacoboni M., Molnar-Szakacs I., Gallese V., Buccino G., Mazziotta JC., Rizzolatti G. (2005) Grasping the intentions of others with one's own mirror neuron system PLoS Boil. 3, e79
- Kaplan JT, Iacoboni M (2007) Multimodal action representation in human left ventral premotor cortex. Cogn Process 8:103–113
- Keysers C, Kohler E, Umilta MA, Nanetti L, Fogassi L et al (2003) Audiovisual mirror neurons and action recognition. Exp Brain Res 153:628–636

- Knoblich G, Repp BH (2009) Inferring agency from sound. Cognition 111:248–262
- Kohler E, Keysers C, Umilta` MA, Fogassi L, Gallese V et al (2002) Hearing sounds, understanding actions: action representation in mirror neurons. Science 297:846–848
- Lang HA, Nyrke T, Ek M, Aaltonen O, Raimo I, Näätänen R (1990) Pitch discrimination performance and auditory event-related potentials. In: Brunia CHM, Gaillard AWK, Kok A, Mulder G, Verbaten MN, editors. Psychophysiological brain research; vol. 1. Tilburg, the Netherlands: Tilburg University Press; 1990. p. 294–8
- Lepage JF, Tremblay S, Nguyen DK, Champoux F, Lassonde M, Théoret H (2010) Action related sounds induce early and late modulations of motor cortex activity. Neuroreport 21:250–253
- Leslie KR, Johnson-Frey S, Grafton ST (2004) Functional imaging of face and hand imitation: towards a motor theory of empathy. NeuroImage 21:601–607
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. J Neurosci 25:5148–5158
- Lewis JW, Phinney RE, Brefczynski-Lewis JA, DeYoe EA (2006) Lefties get it ‘‘right’’ when hearing tool sounds. J Cogn Neurosci 18:1313–1314
- Micheyl, C. *et al.* (2003) The neural basis of stream segregation in the primary auditory cortex. Assoc. Res. Otalryngol. Abs. 26, 195
- Miller, G.A., Heise, G.A. (1950) The trill threshold. J. Acoust. Soc. Am. 22, 637-638
- Molholm S, Martinez A, Ritter W, Javitt D.C., Foxe J.J. (2005) The neural circuitry of Pre-attentive auditory change-detection: An fMRI study of pitch and duration mismatch negativity generators. Cereb. Cortex; 545-551 15(5)
- Moro, V., Urgesi, C., Pernigo, S., Lanteri, P., Pazzaglia, M. & Aglioti, S.M. (2008) The neural basis of body form and body action agnosia. Neuron, 60, 235–246
- Müller D, Widman A, Schröger E (2005) Auditory streaming affects the processing of successive deviant and standard sounds. Psychophysiology; 42:668–76
- Näätänen R (1979) Orienting and Evoked Potentials. In: Kimmel HD, van Olst EH, Orlebeke JF, editors. The orienting reflex in humans. New Jersey: Erlbaum; p. 61–75

- Näätänen R (1985) Selective attention and stimulus processing: reflections in event-related potentials, magnetoencephalogram, and regional cerebral blood flow. In: Posner MI, Marin OSM, editors. Attention and performance 1985;vol. XI. Hillsdale, NJ: Erlbaum. p. 355–73

- Näätänen R, Pakarinen S, Rinne T, Takegata R (2004) The mismatch negativity (MMN) - towards the optimal paradigm. Clin Neurophysiol; 115:140–4

- Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001) ‘‘Primitive intelligence’’ in the auditory cortex. Trends Neurosci; 24:283–8

- Näätänen R, Winkler I (1999) The concept of auditory stimulus representation in neuroscience. Psychol Bull; 125:826–59

- Näätänen R, Winkler I (1999) The concept of auditory stimulus representation in neuroscience. Psychol Bull; 125:826–59

- Paavilainen P, Jiang D, Lavikainen J, Näätänen R (1993a) Stimulus duration and the sensory memory trace: an event-related potential study. Biol Psychol; 35:139–52

- Pakarinen S, Huotilainen M, Näätänen R (2010) The mismatch negativity (MMN) with no standard stimulus. Clin Neurophysiol; 1043:1050-121

- Pakarinen S, Takegata R, Rinne T, Huotilainen M, Näätänen R (2007) Measurement of extensive auditory discrimination profiles using mismatch negativity (MMN) of the auditory event-related potential. Clin Neurophysiol; 118:177–85

- Pazzaglia M, Pizzamiglio L, Pes E, Aglioti SM (2008a) The sound of actions in apraxia. Curr Biol 18:1766–1772 (Epub 2008 Nov 13)

- Pazzaglia M, Smania N, Corato E, Aglioti SM (2008b) Neural underpinnings of gesture discrimination in patients with limb apraxia. J Neurosci 28:3030–3041

- Penfield W, Rasmussen T (1950) The cerebral cortex of man. MacMillan, New York

- Pernigo S., Moro V., Avesani R., Miatello C., Urgesi C., Aglioti SM. (2012) Massive somatic deafferentation and motor deafferentation of the lower part of the body impair its visual recognition: a psychophysical study of patients with spinal cord injury Eur. J. Neurosci., 36, 3509-3518

- Pietrini P, Furey ML, Ricciardi E, Gobbini MI, Wu WH, Cohen L, Guazzelli M, Haxby JV (2004) Beyond sensory images: objectbased representation in the human ventral pathway. Proc Natl Acad Sci USA 101:5658–5663

- Pizzamiglio L, Aprile T, Spitoni G, Pitzalis S, Bates E, D'Amico S, Di Russo F (2005) Separate neural systems for processing action- or non-action-related sounds. Neuroimage 24:852–861

- Pulvermüller F (2001) Brain reflections of words and their meaning. Trends Cogn Sci 5:517–524

- Raos V., Evangeliou MN., SavakiH.E. (2004) Observation of action: grasping with the mind's hand. Neuroimage, 23, 193-201.

- Ricciardi E, Bonino D, Sani L, Vecchi T, Guazzelli M, Haxby JV, Fadiga L, Pietrini P (2009) Do we really need vision? How blind people ‘‘see’’ the actions of others. J Neurosci 29:9719–9724

- Rizzolatti G, Craighero L (2004) The mirror-neuron system. Annu Rev Neurosci 27:169–192

- Rizzolatti G, Fadiga L, Matelli M, Bettinardi V, Paulesu E, Perani D, Fazio F (1996) Localization of grasp representations in humans by PET: 1. Observation versus execution. Exp Brain Res 111:246–252

- Romani M, Cesari P, Urgesi C, Facchini S, Aglioti SM (2005) Motor facilitation of the human cortico-spinal system during observation of bio-mechanically impossible movements. Neuroimage 26:755–763

- Rossi S, De Capua A, Pasqualetti P, Ulivelli M, Fadiga L, Falzarano V, Bartalini S, Passero S, Nuti D, Rossini PM (2008) Distinct olfactory cross-modal effects on the human motor system. PLoS One 27(3):e1702

- Rossi S., Tecchio F., Pasqualetti P., Ulivelli M., Pizzella V., Romani GL., Passero S., Battistini N., Rossini P.M. (2002) Somatosensory processing during movement observation in humans. Clin. Neurophysiol., 113, 16-24

- Saarinen J, Paavilainen P, Schröger E, Tervaniemi M & Näätänen R. (1992) Representation of abstract attributes of auditory stimuli in the human brain. Neuroreport, 3, 1149–1151

- Sakamoto M., Muraoka T., Mizuguchi N., Kanosue K., (2009) Combining observation and imagery of an action enhances human corticospinal excitability. Neurosci. Res. 65, 23-27

- Schubotz RI, von Cramon DY, Lohmann G (2003) Auditory what, where, and when: a sensory somatotopy in lateral premotor cortex. Neuroimage 20:173–185

- Serino, A., De Filippo, L., Casavecchia, C., Coccia, M., Shiffrar, M. & Ládavas, E. (2010) Lesions to the motor system affect action perception. J. Cogn. Neurosci., 22, 413–426
- Stefan K, Cohen LG, Duque J, Mazzocchio R, Celnik P, Sawaki L, Ungerleider L, Classen J (2005) Formation of a motor memory by action observation. J Neurosci 25:9339–9346
- Strafella AP, Paus T (2000) Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. Neuroreport 11:2289–2292
- Sussman E, Ritter W, Vaughan Jr HG (1998) Attention affects the organization of auditory input associated with the mismatch negativity system. Brain Res; 789:130-8
- Sussman E., Ritter W., &Vaughan H. G. Jr. (1999). An investigation of the auditory streaming effect using event-related brain potentials. Psychophysiology, 36, 22–34
- Takada M, Tokuno H, Nambu A, Inase M (1998) Corticostriatal projections from the somatic motor areas of the frontal cortex in the macaque monkey: segregation versus overlap of input zones from the primary motor cortex, the supplementary motor area, and the premotor cortex. Exp Brain Res 120:114–128
- Tettamanti M, Buccino G, Saccuman MC, Gallese V, Danna M et al (2005) Listening to action-related sentences activates frontoparietal motor circuits. J Cogn Neurosci 17:273–281
- Tiitinen H, Alho K, Huotilainen M, Ilmoniemi RJ, Simola J, Näätänen R (1993) Tonotopic auditory cortex and the magnetoencephalographic (MEG) equivalent of the mismatch negativity. Psychophysiology; 30:537–40
- Tiitinen H, May P, Reinikainen K, Näätänen R (1994) Attentive novelty detection in humans is governed by pre-attentive sensory memory. Nature 1994;370:90–2
- Tokuno H, Takada M, Nambu A, Inase M (1997) Reevaluation of ipsilateral corticocortical inputs to the orofacial region of the primary motor cortex in the macaque monkey. J Comp Neurol 389:34–48
- Urgesi C, Candidi M, Fabbro F, Romani M, Aglioti SM (2006) Motor facilitation during action observation: topographic mapping of the target muscle and influence of the onlooker’s posture. Eur J Neurosci 23:2522–2530
- Urgesi C, Maieron M, Avenanti A, Tidoni E, Fabbro F, Aglioti SM (2010) Simulating the future of actions in the human corticospinal system. Cereb Cortex

- Urgesi, C., Berlucchi, G. & Aglioti, S.M. (2004) Magnetic stimulation of extrastriate body area impairs visual processing of nonfacial body parts. Curr. Biol., 14, 2130–2134

- Urgesi, C., Calvo-Merino, B., Haggard, P. & Aglioti, S.M. (2007a) Transcranial magnetic stimulation reveals two cortical pathways for visual body processing. J. Neurosci., 27, 8023–8030

- Urgesi, C., Candidi, M., Ionta, S. & Aglioti, S.M. (2007b) Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. Nat. Neurosci., 10, 30–31

- Van Noorden, L.P.A.S. (1975) Temporal Coherence in the Perception of Tone Sequences. Ph.D. Thesis, Tech. Hogeschool, Eindhoven; Institute of Perception Research, The Netherlands

- van Zuijen TL *et al.* (2005) Auditory organization of sound sequences by a temporal or numerical regularity: a mismatch negativity study comparing musicians and non-musicians. Cogn. Brain Res. 23, 270–276

- Vogt S, Buccino G, Wohlschla¨ger AM, Canessa N, Shah NJ, Zilles K, Eickhoff SB, Freund HJ, Rizzolatti G, Fink GR (2007) Prefrontal involvement in imitation learning of hand actions: effects of practice and expertise. Neuroimage 37:1371–1383

- Waters, R.L., Adkins, R.H. & Yakura, J.S. (1991) Definition of complete spinal cord injury. Paraplegia, 29, 573–581

- Wheaton KJ, Thompson JC, Syngeniotis A, Abbott DF, Puce A (2004) Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. NeuroImage 22:277–288

- Wilson M., Knoblich G., (2005) The case for motor involvement in perceiving conspecifics. Psychol. Bull., 131, 460-473

- Winkler I, Cowan N, Cse´pe V, Czigler I, Näätänen R (1996a) Interactions between transient and long-term auditory memory as reflected by the mismatch negativity. J Cogn Neurosci; 8:403–15

- Winkler I, Karmos G, Näätänen R (1996b) Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. Brain Res; 742:239–52

- Winkler I, Kujala T, Tiitinen H, Sivonen P, Alku P, Lehtokoski A, *et al.* (1999b) Brain responses reveal the learning of foreign language phonemes. Psychophysiology; 36:638–42

- Winkler I, Lehtokoski A, Alku P, Vainio M, Czigler I, Csépe V, *et al.* (1999a) Preattentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. Cogn Brain Res;7:357–69
- Winkler I, Schröger E, Cowan N. (2001) The role of large-scale perceptual organization in the mismatch negativity event-related brain potential. J Cogn Neurosci; 13:59–71
- Winkler, I. (2007) Interpreting the mismatch negativity. J. Psychophysiol. 21, 147–163
- Yost, W.A. (1991) Auditory image perception and analysis: the basis for hearing. Hear. Res. 56, 8-18
- Zachau S., Rinker T., Körner B., Kohls, G., Maas V., Hennighausen K., Schecker M. (2005) Extracting rules: early and late mismatch negativity to tone patterns. Neuroreport 16, 2015-2019

APPENDIX A

Overview of Publication Status of Chapters in Thesis

Chapter number and title	Original text (not published before)	Submitted: no feedback received	Submitted: revision requested or revision submitted	Accepted/published (specify journal or book)
1. Introduction	X			
2. Chapter One: **Mechanisms and neural underpinnings of audio-motor action mapping**	X			
3. Chapter Two: **Deviance detection and Auditory Scene Analysis (ASA)**	X			
4. Chapter Three: **Investigating the stream segregation in audio-motor mapping**	X In preparation			
5. Chapter Four: **The role of massive somatic deafferentation and motor deefferentation of the lower and upper part of the body in audio-motor mapping**	X In preparation			
6. **General Discussion**	X			