EPFL

Dissecting self-voice perception: From bone conduction to robotically-induced self-other voice misattribution in healthy listeners

Pavo OREPIC

■ École
polytechnique
fédérale
de Lausanne

Za dida Pavu, koji me naučio prva slova.

For my grandpa Pavo, who taught me my first letters.

Acknowledgements

First and foremost, I would like to sincerely thank my thesis director – Prof. Olaf Blanke – for being such an outstanding supervisor, a great source of inspiration and for teaching me how to think big. Somehow, he always managed find just the right words and after every meeting with him, I would end up being even more motivated. From the first day, I could feel that he believed in me and I thank him for giving me a lot of freedom to implement all my ideas. I am very grateful for having an opportunity to start my academic career with him.

I also need to thank Nathan, who was closely supervising me in the first part of my PhD, before starting his own lab. I thank him for his scientific rigor, for always finding time for me, and for always having an answer for every single question I made – be it a how-to in R, or giving me the right reference by heart.

Through different projects, I was fortunate to work with most of the postdoc of LNCO, and I thank them for all the help, amazing supervision, and challenging questions. A special thanks to Olli and Fosco for their availability, especially in the final stages of my thesis.

I also had the opportunity to spend the first six months of my PhD in the USA with Prof. Ali Rezai, a world leader in neuromodulation and fantastic team manager. I am grateful for his time, advice and extraordinary hospitality. As a young computer scientist, starting my neuroscience PhD by working with rare cohorts of neurosurgical patients was a tremendous source of inspiration.

A big thanks also goes to Giannarita, a collaborating postdoc at UNIGE, with who I worked on my EEG and clinical projects at HUG. She introduced me to the world of EEG and personally supervised me in doing advanced analysis, step by step. I thank her for all the patience, determination, availability and many funny moments.

I am very grateful to Jelena, the director of SFITS, who became my mother in Switzerland, basically from the minute she met me. I thank her for taking a good care of me, making sure I ate, and for all many great moments made in such a short time.

I would like to thank my family, for their unconditional love and support and for still not giving up on trying to understand what is it exactly that I do. A special thanks to my dad, my best friend, to whom I owe everthing that I achieved. He closely followed every step I made, making sure that I would be able to make it, no matter what. In fact, so closely that, even today, he gets angry if we don't talk for three days.

A huge thanks to my second family – Čići, Marina, Pijer, Melo, Vaba, Mia, and Tomo – my childhood friends, who are a living proof that distance is just a myth.

I also need to thank Matej, who I met during my studies in Germany and with whom I became unseparable from the moment we met. I thank him for knowing where I was, even at times when I did not.

My life in Switzerland would not have been the same without Herberto, Giuliana, Matteo, Giedre, Nathalie, and Eva. I was so lucky to get to spend these 4 years with them and to share countless wonderful moments. We have been riding the rolloercoaster of science together, with all the ups and downs, and always managed to make it an exciting adventure.

Finally, the biggest thanks goes to Julia, without whom none of this would have been possible. She was with me through all the best and all the worst, somehow providing me with what I needed before I knew I needed it. She also showed me that it is possible to eat food that does not grow in Croatia. Generally, she made me a better version of myself. To her, I just thank for being.

Lausanne, 15 September 2020

Pavo

Abstract

Inspired by the prominent theory suggesting that auditory-verbal hallucinations (AVH) – the sensation of hearing voices without present speakers – arise as misattribution of inner speech towards external agents, my thesis revolved around self-voice perception and the experimental attempts of self-other voice misattribution in a healthy, non-hallucinating population, potentially mimicking the AVH phenomenology.

In the first part of my thesis, I investigated behavioral (Study 1) and neural (Study 2) underpinnings of self-other voice discrimination (SOVD). Compared to other self-related processes, self-voice perception has been investigated to a surprisingly lesser extent. Namely, self-voice research has been thwarted by the inability to experimentally match self-voice recordings to the natural sound of our voice that is altered by bone conduction. In a series of five experiments, I showed that this discrepancy can be reduced by presenting self-voice stimuli through a commercial bone conduction headset, thereby rendering self-voice as an essentially multimodal construct. My data further shows that self-voice recognition differs from the recognition of familiar voices, however, that it still involves some familiarity processing. In addition, I explored the roles other-voice familiarity and acoustic similarity to other voice play in SOVD. Finally, I identified a self-voice specific EEG pattern, around 345 milliseconds after stimulus onset, that followed the initial auditory cortex activation, discriminated the self-voice from the voice of another unfamiliar person, and activated an extended network involving the cingulate cortex, insula, and medial temporal lobe structures. Moreover, this network was recruited less frequently with self-voices presented through bone conduction, and the occurrence of the network negatively correlated with SOVD task performance.

Based on these methodological, behavioural, and neural findings in SOVD, in the second part of my thesis, I tried to experimentally induce specific misattributions of self-towards-other voices, following the prominent theory linking AVH with self-monitoring deficits. In order to do so, I combined the aforementioned SOVD task with a robotic procedure able to engender mild hallucinations in healthy individuals by perturbing bodily self-monitoring mechanisms. With such a procedure, I managed to alter self-voice perception of healthy participants (Study 3), what I further related to breathing (Study 4). Finally, I managed to induce identity-specific AVH in healthy individuals (Study 5), as quantified by the false alarm rate in a voice detection task, thereby contributing to the understanding of the mechanisms underlying AVH and relating them to impairments in bodily self-monitoring.

Laying at the intersection of three sub-fields of neuroscience – voice perception, self-processing, and psychiatry – my thesis made a contribution by identifying neural correlates of SOVD, by proposing a new, multimodal perspective on self-voice-related research questions, and by demonstrating an experimental method of inducing AVH in a controlled laboratory environment. Together, my findings shed new light on the interactions between self-voice perception, sensorimotor processing, and interoception.

Keywords

Self-voice, self-other voice discrimination, bone conduction, auditory-verbal hallucinations, self-monitoring, sensorimotor conflicts, psychophysics, EEG, insula, breathing

Résumé

Inspirée par la théorie éminente suggérant que les hallucinations auditives-verbales (HAV) – ou la sensation d'entendre des voix sans locuteur présent - sont le résultat d'une attribution erronée de d'un discours intérieur à des agents extérieurs, ma thèse a tourné autour de la perception de la voix propre et de tentatives expérimentales d'attribution erronée de la voix propre à autrui dans une population saine et non hallucinante, imitant potentiellement la phénoménologie des HAV.

Dans la première partie de ma thèse, j'ai étudié les fondements comportementaux (étude 1) et neuronaux (étude 2) de la discrimination la voix propre de la voix d'autrui (self-other voice discrimination, SOVD). Comparée à d'autres processus liés au soi, la perception de la voix d'autrui a été étudiée dans une mesure étonnamment moindre. En effet, la recherche sur la voix propre a été contrecarrée par l'impossibilité de faire correspondre expérimentalement les enregistrements de la voix propre au son naturel de notre voix qui est altéré par la conduction osseuse. Dans une série de cinq expériences, j'ai montré qu'il était possible de réduire cet écart en présentant les stimuli de la voix propre par le biais d'un casque à conduction osseuse commercial, rendant ainsi la transmission de la voix propre multimodale. Mes données montrent en outre que la reconnaissance de la voix propre diffère de la reconnaissance des voix familières, mais qu'elle implique tout de même un certain traitement de familiarité. En outre, j'ai investigué les rôles que jouent la familiarité et la similarité acoustique d'une autre voix dans la SOVD. Enfin, j'ai identifié un modèle EEG spécifique à la voix propre environ 345 millisecondes après le début du stimulus. Il suit l'activation initiale du cortex auditif, permet la distinction entre la voix propre et la voix d'une personne inconnue et il active un réseau étendu comprenant le cortex cingulaire, l'insula et les structures du lobe temporal médian. De plus, ce réseau est moins souvent recruté lors de la présentation de la voix propre par conduction osseuse, et l'apparition du réseau était négativement corrélée avec la performance d'une tâche SOVD.

Sur la base de ces résultats méthodologiques, comportementaux et neuronaux de la SOVD, j'ai tenté, dans la deuxième partie de ma thèse, d'induire expérimentalement des erreurs d'attribution de la voix propre vers la voix d'autrui la théorie proéminente qui lie l'HAV à des déficits d'autosurveillance. Pour ce faire, j'ai combiné la tâche SOVD susmentionnée avec une procédure robotique capable d'engendrer de légères hallucinations chez des individus en bonne santé en perturbant les mécanismes d'autosurveillance corporelle. Grâce à cette procédure, j'ai réussi à modifier la perception de la voix des participants en bonne santé (étude 3), ce que j'ai ensuite relié à la respiration (étude 4). Enfin, j'ai réussi à induire une HAV spécifique à l'identité chez des personnes en bonne santé (étude 5), quantifiée par le taux de fausses alertes dans une tâche de détection vocale, contribuant ainsi à la compréhension des mécanismes sous-jacents de l'HAV, en les reliant aux troubles de l'autosurveillance corporelle.

Située à l'intersection de trois sous-domaines des neurosciences - la perception de la voix propre, le traitement du soi et la psychiatrie - ma thèse a apporté une contribution en identifiant les corrélats neuronaux de la SOVD, en proposant une nouvelle perspective multimodale sur les questions de recherche liées à la voix propre et en démontrant une méthode expérimentale d'induction de l'HAV dans un environnement de laboratoire contrôlé. Ensemble, mes résultats jettent un nouvel éclairage sur les interactions entre la perception de la voix propre, le traitement sensorimoteur et l'intéroception.

Mots-clés

Voix propre, discrimination de la voix propre de la voix d'autrui, conduction osseuse, hallucinations auditives-verbales, autosurveillance, conflits sensorimoteurs, psychophysique, EEG, insula, respiration

Contents

List of Figures

List of Tables

Introduction

> ❝ *Perceiving your own voice means perceiving your true self or nature. When you and the sound become one, you don't hear the sound; you are the sound.*

> Seung Sahn

For something so central to our identities, it is surprising how little is known about self-voice perception. How do I know that the voice I hear is mine? Is recognizing my voice the same as recognizing the voice of a close friend or a family member? How does the brain discriminate a stranger's voice from my own? How I am I sure that the voice I just heard was real? The reason for the underrepresentation of self-voice research, compared to extensive work on self-face or self-body representations, is mainly methodological. Namely, when we speak, we hear our own voice both through air and bone conduction, and previous experiments have exclusively employed air conduction to investigate self-voice perception, causing a discrepancy between experimental self-voice stimuli and the sound of one's own voice heard during natural speech. However, acquiring a better understanding of self-voice perception is of utmost importance, as erroneous self-voice misattribution has been proposed to account for auditory-verbal hallucinations (AVH) (Frith, 1992; Frith & Done, 1989), the most common hallucination in schizophrenia (Bauer et al., 2011), associated with high degrees of distress in the affected population (Harkavy-Friedman et al., 2003).

The main research aims of my thesis were twofold. For one, I investigated behavioral and neural mechanisms underlying self-voice perception by combining psychophysics, voice-morphing technology (Kawahara, Morise, Banno, & Skuk, 2013), and high-density EEG to develop a sensitive measure of self-other voice discrimination. I studied the effects of bone conduction on self-voice perception, thereby trying to overcome the methodological limitations present in contemporary self-voice studies. I further explored the relationship between the ability to recognize self-voice and autonomic functions such as breathing. For another, I developed experimental procedures aimed at inducing AVH in healthy individuals in a controlled laboratory environment, and thus contributed the better understanding of its underlying principles. In order to do so, I extended the robotic procedure, previously related to mild psychosis-like sensations (Blanke et al., 2014; Salomon et al., 2020; Faivre et al., 2020; Bernasconi, Blondiaux et al., 2020), with precise self-voice manipulations that are thought to account for voice-hearing sensations in schizophrenia (Frith, 1992; Frith & Done, 1989).

In the next sections, I will introduce the general background needed for a good understanding of the present thesis. I will start by defining bodily self-consciousness (BSC) and its dependence on multisensory and sensorimotor integration, thereby introducing the aforementioned robotic procedure. Then, I will narrow BSC down to self-voice perception by providing an overview of the state of the art and challenges self-voice research is facing. I will finish by presenting the theoretical framework relating deficits in self-voice perception to AVH.

1.1   Bodily self-consciousness

Being one of the most astonishing features of the human mind, self-consciousness is investigated in a wide range of disciplines including philosophy, psychology, psychiatry, sociology and cognitive neuroscience. Next to high-order cognitive aspects of self-consciousness (e.g., related to memory, language, or social identity), the sense of self arguably implies the pre-reflective conscious experience of being a self inside a body, i.e. the "minimal self", denoted as bodily self-consciousness (Blanke & Metzinger, 2009; Blanke, Slater, & Serino, 2015; Gallagher, 2000; Jeannerod, 2003). BSC entails the feeling of being localized in one's own body (self-location), experiencing it as one's own (self-identification) and experiencing to take the perspective from that body (Blanke, 2012; Blanke & Metzinger, 2009; Blanke et al., 2015).

1.1.1   Multisensory and sensorimotor integration

BSC is built upon a continuous integration of diverse exteroceptive (e.g. visual, auditory, or tactile) as well as interoceptive (e.g. visceral, respiration, or heartbeat) signals (Blanke, 2012; Blanke, Slater, & Serino, 2015; Park & Blanke, 2019; Tsakiris, 2017). This is corroborated by experimental evidence demonstrating that multisensory integration of conflicting bodily signals leads to altered states of BSC, such as changes in self-identification, self-location and first-person perspective (Blanke & Metzinger, 2009; Lenggenhager, Tadi, Metzinger, & Blanke, 2007; Pfeiffer, Schmutz, & Blanke, 2014; Preston, Kuper-Smith, & Ehrsson, 2016). Interestingly, it has been shown that BSC can be altered even if one of the conflicting bodily signals is interoceptive (Adler, Herbelin, Similowski, & Blanke, 2014; Aspell et al., 2013; Heydrich et al., 2018; Suzuki, Garfinkel, Critchley, & Seth, 2013; for a review see Park & Blanke, 2019) or unconscious (Salomon et al., 2017). Experimental manipulations of BSC have moreover been shown to affect higher cognitive processing (Canzoneri, Herbelin, Blanke, & Serino, 2016) and have been extended to clinical applications involving stroke, chronic pain and pathological bodily images (Nierula, Martini, Matamala-Gomez, Slater, & Sanchez-Vives, 2017; Pozeg et al., 2017; Ronchi, Heydrich, Serino, & Blanke, 2017; Scandola, Aglioti, Pozeg, Avesani, & Moro, 2017; Solcà et al., 2018).

Closely related to BSC is the sense of agency (SoA) – the feeling of being the one agent causing or generating an action and its sensory consequences (Gallagher, 2000). As opposed to multisensory integration accounting for the aspects of BSC such as self-identification and self-location, SoA arises from sensorimotor integration related to the interactions of the body with the environment (Braun et al., 2018; O. A. Kannape & Blanke, 2012; Tsakiris & Haggard, 2005), building up on the mechanisms of motor control (Jeannerod, 2006; Shadmehr, Smith, & Krakauer, 2010; Wolpert, 1997). SoA allows to dissociate events that are self-generated from those caused by the environment (Jeannerod, 2003) or another agent (Jeannerod & Pacherie, 2004), which is necessary for constructing and maintaining a coherent and stable representation of the self. According to the prominent self-monitoring framework (Blakemore, Wolpert, & Frith, 2000; Miall & Wolpert, 1996; Wolpert, Ghahramani, & Jordan, 1995), this is achieved by creat-
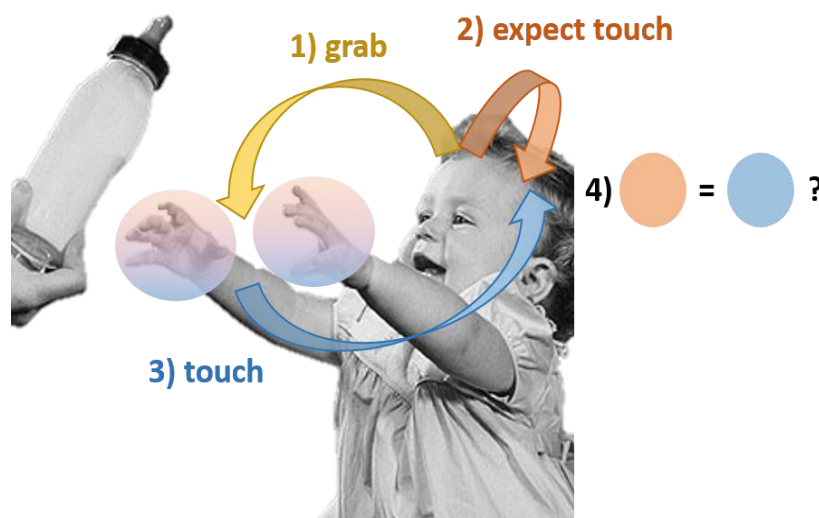
Figure 1. Illustration of the self-monitoring framework
Motor action (1, yellow) is followed by an expectation of the sensory feedback (2, orange) that might result from the action. After the action is performed, the actual sensory feedback (3, blue) is compared to the expected one (4). If there is a match between the expected and the actual feedback, neural processing of the feedback is attenuated and the feedback (e.g. touch) is attributed to oneself. In case of a mismatch, the feedback is attributed to someone other than self. *Image adapted from Getty Images.*

ing sensory predictions of own actions and by comparing them with the actual sensory feedback following those actions. When congruent with the prediction, sensory feedback is attenuated, and the action is attributed to the self, whereas if incongruent, there is no attenuation and the action is attributed to another agent (Figure 1). Consequently, self-generated sensations are perceived as less salient and exert smaller neural responses compared to externally generated ones. First pondered upon in visual perception (Helmholtz, 1866), self-monitoring was further observed in somatosensation (Blakemore et al., 2000; Blakemore, Wolpert, & Frith, 1998; Shergill, Bays, Frith, & Wolpert, 2003) and audition (Greenlee et al., 2011; Martikainen, Kaneko, & Hari, 2005; Paus, Perry, Zatorre, Worsley, & Evans, 1996; Schafer & Marcus, 1973).

1.1.2   Robotically-mediated sensorimotor stimulation

Until recently, SoA has mostly been investigated for upper-limb actions (e.g. pressing a button) and experimentally manipulated with robotic devices that facilitate sensorimotor conflicts between hand movements and the corresponding sensory feedback on the hand (Blakemore, Frith, & Wolpert, 1999; Shergill et al., 2003; Weiskrantz, Elliott, & Darlington, 1971). However, as research on BSC has demonstrated the importance of representing the self as a single spatially-situated global unit, rather than separated body parts (Blanke & Metzinger, 2009), the concept of agency has been extended to movements of the body as a whole (e.g. gait; Kannape and Blanke, 2013, 2012; Menzer et al., 2010). Accordingly, (Hara et al., 2011) designed a robotic device (Figure 2., left) that associated upper-limb sensory prediction signals with reafferent sensory signals at the participants' torso in order to alter the representation of the global, torso-centered bodily system (Blanke & Metzinger, 2009; Park & Blanke, 2019).

Using this robotic device, (Blanke et al., 2014) were able to induce systematic changes in illusory own body perceptions (i.e. self-touch) and mild psychosis-like phenomena, in healthy volunteers, that depended on sensorimotor conflicts (Figure 2, right). Specifically, while perceiving spatiotemporal sensorimotor conflicts between poking movements and touches on their backs, participants reported stronger somatic passivity (i.e. that tactile sensations are being imposed upon their body by somebody else) and felt being in a presence of a non-existing alien entity, phenomenologically resembling passivity experiences (Frith, Blakemore, & Wolpert, 2000; Sass & Parnas, 2003; Sass & Parnas, 2001) and presence hallucinations (Alderson-Day & Fernyhough, 2016; Critchley, 1955; Jaspers, 1990) observed in schizophrenia. Thus, more than just a loss of agency that typically occurs with upper-limb sensorimotor conflicts, torso-oriented sensorimotor stimulation is able to introduce other-agency sensations, thereby mirroring the related symptomology in schizophrenia.
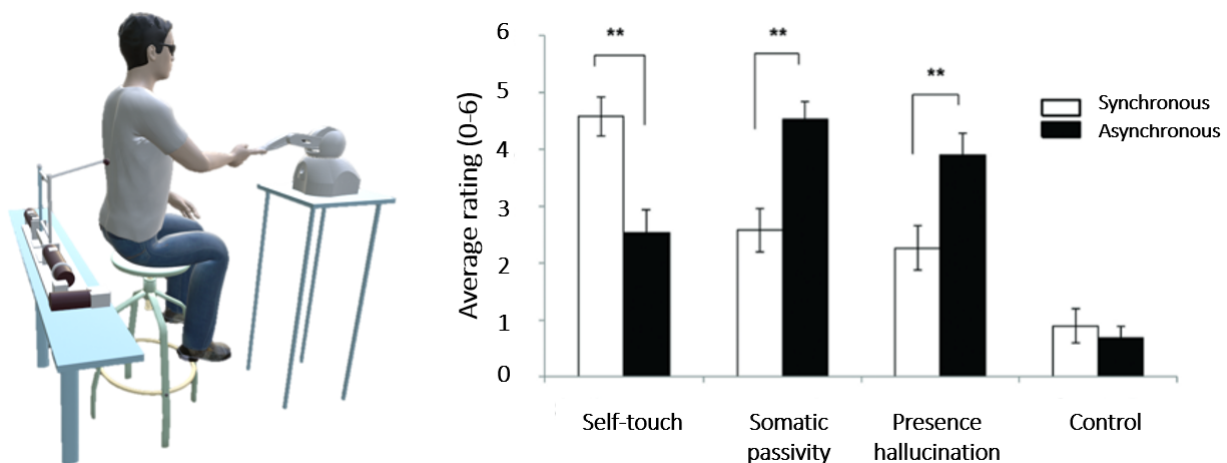
Figure 2. Robotically-induced mild psychosis-like states
Left: Robotic procedure able to induce a mild psychosis-like state in a safe and controlled manner in healthy individuals. Blindfolded participants are performing repetitive forward-backward poking movements on the front part of the robot, which is replicated by the back part of the robot in a synchronous or asynchronous (i.e. with a delay) fashion. Right: Likert-scale ratings assessing subjective experience following two minutes of robotic sensorimotor stimulation. Following synchronous stimulation, which constitutes spatial sensorimotor conflict between the pokes and the corresponding touches, participants report stronger self-touch sensations. After asynchronous stimulation, which has an additional temporal sensorimotor conflict, participants experience stronger somatic passivity and presence hallucination. Adapted from (Blanke et al., 2014). **: $p<0.01$

1.2 Self-voice perception

Besides bodily signals, there is growing evidence suggesting that any self-related stimulus is processed as special, i.e. different from other (highly) familiar stimuli (Northoff et al., 2006; Qin, Wang, & Northoff, 2020; Sui & Humphreys, 2017) (although some disagree (Gillihan & Farah, 2005)). The two stimuli most associated with a person's identity are her face and voice (Blank, Wieland, & von Kriegstein, 2014). Although it has been suggested that the process of person identification is similar in both modalities (Yovel & Belin, 2013), it is easier to recognize a familiar face, compared to a familiar voice (Hanley, Smith, & Hadfield, 1998), and seeing a familiar face allows for retrieving more biographical information about the person (Damjanovic & Hanley, 2007). What about own faces and voices? Self-face representation has been extensively investigated (for an overview see Uddin, Kaplan, Molnar-Szakacs, Zaidel, & Iacoboni, 2005), and there is considerable evidence indicating the special nature of the self in visual modality (Alzueta, Melcón, Poch, & Capilla, 2019). Self-voice, however, has been investigated to a surprisingly lesser extent, and no such claim could yet be made.

1.2.1 Neural correlates of hearing self-voice

To date, only a few neuroimaging studies contrasted activations following self-voice and other-voice presentation. (Allen et al., 2005) associated self-voice to activity in right anterior cingulate and left inferior frontal cortex. In the PET study of (Nakamura et al., 2001), a contrast between self-voice and familiar voice activation peaked in the right inferior frontal sulcus and parainsular cortex. Similarly, right inferior frontal gyrus produced greater signal to self- compared to familiar voice in a study of (Kaplan, Aziz-Zadeh, Uddin, & Iacoboni, 2008).

EEG investigations of self-voice specificity have been confined to the analysis of single-electrode evoked responses following self and other voices. Most studies contained an oddball paradigm and reported a lower P3 component fronto-central sites (FCz) for self-voice, compared to unfamiliar (Graux et al., 2013) and familiar voices (Graux, Gomot, Roux, Bonnet-Brilhault, & Bruneau, 2015). However, slightly different experimental designs yielded in contradictory results (Liu, Li, Li, Lou, & Chen, 2019), sometimes even within the same group of researchers (Conde, Gonçalves, & Pinheiro, 2016, 2018). Although inconsistently, these data suggest that self-voice is represented differently from other voices.

1.2.2 Importance of bone conduction

These findings, however, should be taken with caution due to the methodological limitation present throughout all self-voice studies – the stimuli consistently presented in these studies (own voice recordings) do not correspond to our internal self-voice representation as based on ecological audio conditions – i.e. they do not sound natural to us. Interestingly, many studies that compared recognition of own vs other voices in a behavioral task showed lower accuracy or response times for self-voice stimuli (Allen et al., 2005; Gur & Sackeim, 1979; Hughes & Nicholson, 2010; Rosa, Lassonde, Pinard, Keenan, & Belin, 2008; Schuerman, Meyer, & McQueen, 2015; Shuster, 1998), indicating inability to recognize the self-voice recordings as well as recordings of other voices. Moreover, hearing our voice in a voice recording is very often accompanied with feelings of unpleasantness (Sackeim & Gur, 1978), related to the mismatch between the sound of own voice in recordings and the sound of our voice we are accustomed to while speaking. This discrepancy arises due to lack of bone conduction, which is inevitably present while speaking (Békésy, 1949; Reinfeldt, Östli, Håkansson, & Stenfelt, 2010). Namely, while speaking, we hear ourselves not only through air, but also through bone conduction (Reinfeldt et al., 2010). Although extensively investigated (Dobrev et al., 2017; Pörschmann, 2000; Stenfelt, 2011; Stenfelt & Goode, 2005), the exact transfer function describing bone-conduction-related transformations of our voice still remains unidentified (Stenfelt, 2016). However it is believed to represent a low-pass filter (Tonndorf, 1976; Wheatstone, 1827), meaning that while speaking our voice sounds deeper to us than to other people (and accordingly deeper than in own-voice recordings). Interestingly, (Shuster & Durrant, 2003) have shown that low-pass filtered sound of our voice sounds more natural to us compared to regular own-voice recordings, however, more systematic approaches did not come to a similar conclusion (Kimura & Yotsumoto, 2018; Maurer & Landis, 1990).

In addition to the physical transformation of the sound of our voice, during speech our voice is also inevitably accompanied by a vibrotactile excitation (Stenfelt, 2011) caused by bone conduction of the sound through the skull. Namely, when we speak, not only auditory, but also vestibular (Emami et al., 2012; Todd, Cody, & Banks, 2000) and somatosensory (Ito, Tiede, & Ostry, 2009; Tremblay, Shiller, & Ostry, 2003) processes are often involved. It is therefore possible that in order to increase familiarity to self-voice recordings, not only the physical properties of the sound of our voice should be approximated to the experience of speaking, but also the stimulus presentation involving somatosensory and vestibular excitation. A multimodal presentation of own voice can be achieved through commercial bone conduction headsets that have recently become commercially available and one study has

even demonstrated that they enable better speech recognition compared to regular air-conduction headsets (Manning, Mermagen, & Scharine, 2017).

1.3 Auditory-verbal hallucinations

Auditory-verbal hallucinations (AVH) – the sensation of hearing voices without speakers present – are the most common symptom of psychosis, affecting more than 70% of people suffering from schizophrenia-spectrum disorders (Bauer et al., 2011; Nayani & David, 1996; Sartorius et al., 1986). Being predominantly negative in content and in affect, AVH represent a major source of distress in patients with psychosis and their presence alone significantly increases risk of suicide in this group (Harkavy-Friedman et al., 2003; Kelleher et al., 2012; Pompili et al., 2007). Although antipsychotic medications often succeed in ameliorating hallucinations, AVH prove persistent in 10 to 30% of sufferers (Lehman et al., 2004). Moreover, the phenomenological experience of AVH is heavily heterogeneous (e.g. with respect to voice numerosity, gender, frequency, emotional affect, etc.) (McCarthy-Jones et al., 2014; Woods, Jones, Alderson-Day, Callard, & Fernyhough, 2015) and they have as well been observed in non-clinical populations (Daalman, Diederen, Hoekema, Van Lutterveld, & Sommer, 2016; Johns et al., 2004; Powers, Mathys, & Corlett, 2017; Powers, Kelley, & Corlett, 2017; Sommer et al., 2010), blurring their status as diagnostic markers for mental illness. Thus, understanding the etiology of AVH is a critical next step towards the development of new diagnostic tools and treatments that are more soundly based upon systems neuroscience and brain pathophysiology. There are two prominent and seemingly opposed perspectives on how AVH might arise: 1) deficits in self-monitoring; and 2) precision-weighting of perceptual priors. Although both of these carry some empirical support, only theoretical speculations (Corlett et al., 2019; Northoff & Qin, 2011; Swiney & Sousa, 2014; Synofzik & Vosgerau, 2012; Synofzik, Vosgerau, & Newen, 2008; Wilkinson, 2014) have been made on how they might coexist in the brain and relate phenomenologically.

1.3.1 Self-monitoring impairments

The first account for AVH suggests that they arise as a deficit in self-monitoring mechanisms (Figure 1) Deficits in self-monitoring have been observed in schizophrenia (Blakemore, Smith, Steel, Johnstone, & Frith, 2000; Shergill, Samson, Bays, Frith, & Wolpert, 2005), and, as such, have been related to various psychotic symptoms characterized by a misattribution of self-generated actions towards external agents (Feinberg, 1978; Frith, 1987; Frith, Blakemore, & Wolpert, 2000). Accordingly, some have proposed that AVH may arise as a self-to-other misattribution of inner speech (Ford & Mathalon, 2005; Ford, Roach, Faustman, & Mathalon, 2007; Frith, 1992; Frith & Done, 1988; Frith & Done, 1989; Moseley, Fernyhough, & Ellison, 2013), resulting from erroneous self-monitoring mechanisms associated to speaking (Figure 3).

Despite its theoretical appeal, there is no direct evidence relating inner speech misattribution to the pathogenesis of AVH. This assumption is mainly based on previous reports suggesting that both AVH (Gould, 1948; Green & Kinsbourne, 1990; Green & Preston, 1981; McGuigan, 1966) and inner speech (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Jacobson, 1931; Livesay, Liebke, Samaras, & Stanley, 1996; McGuigan & Dollins, 1989; Wildgruber, Ackermann, Klose, Kardatzki, & Grodd, 1996) may serve as motor actions. However, these reports have never been replicated and the phenomenology of inner speech is still under debate (Hurlburt, Heavey, & Kelsey, 2013; Lœvenbruck et al., 2018; Perrone-Bertolotti, Rapin, Lachaux, Baciu, & Lœvenbruck, 2014). Further empirical support comes from studies (reviewed by (Whitford, 2019)) in which patients with schizophrenia exhibited a reduced suppression of auditory evoked response while speaking compared to passively hearing their voice, what was hypothesized to reflect a self-monitoring deficit. However, such findings were rarely related to hallucinations (Whitford, 2019) and the applied experimental protocols were criticized for lacking important control conditions (Hughes, Desantis, & Waszak, 2013). Then, some researchers observed differences in functional connectivity (Ford, Mathalon, Whitfield, Faustman, & Roth, 2002; Hoffman, Fernandez, Pittman, & Hampson, 2011) in schizophrenia patients, which was, again, assumed to reflect a self-monitoring deficit. Only a few studies (reviewed by (Allen, Aleman, & McGuire, 2007)) directly assessed self-to-other misattribution by playing distorted self- or other-
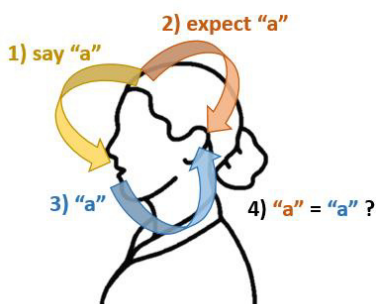
Figure 3. Self-monitoring account for AVH

The intention to speak (1, yellow) is followed by an expectation of the sensory feedback (2, orange) that might result from speaking. After speaking, the actual sensory feedback (3, blue) is compared to the expected one (4). If there is a match between the expected and the actual feedback, neural processing of the feedback is attenuated and the speaker attributes the sound she hears (e.g. "a") to herself. In case of a mismatch, the sound is attributed to someone other than self. It is hypothesized that AVH arise as an error in this process, resulting in a self-to-other misattribution of inner speech.

voice feedback to patients while they spoke aloud. Unfortunately, these attempts yielded equivocal findings, mainly due to methodological limitations and inability to disentangle cognitive impairments from self-monitoring deficits (Allen et al., 2007). Further, reports of own-voice misattributions occurring without having patients speaking (Allen et al., 2004; Mechelli et al., 2007) bring into question its relationship to self-monitoring, which, per definition, necessitates motor action. Finally, self-monitoring deficits alone do not explain the rich phenomenological aspects nor the considerable heterogeneity of AVH (Wilkinson, 2014).

1.3.2 Over-weighting of perceptual priors

Another recent view suggests that hallucinations might be engendered by overly strong prior beliefs (i.e. priors) about the environment. It relies on the predictive coding framework (Adams, Brown, & Friston, 2015; Fletcher & Frith, 2009; Sterzer et al., 2018) that sees the brain as a hierarchical Bayesian system in which priors (higher level) and incoming sensory information (lower level) are combined into perception. More precisely, top-down priors affect perception by conforming the incoming sensory evidence to the existing beliefs. In turn, bottom-up sensory evidence leads to updating of priors, maintaining the stability of those beliefs. Crucially, precision-weighting of bottom-up and top-down components determines the nature of perception, whereby the component with higher precision dominates perception (Siemerkus, Tomiello, & Stephan, 2019). Accordingly, hallucinations have been hypothesized to arise when priors carry undue precision and thus overshadow the actual sensory evidence (Corlett et al., 2019) (Figure 4.). In real-life, strong priors can be observed in several well-known illusions, such as the Hollow Mask Illusion (Gregory, 1973) or the McGurk effect (McGurk & MacDonald, 1976).
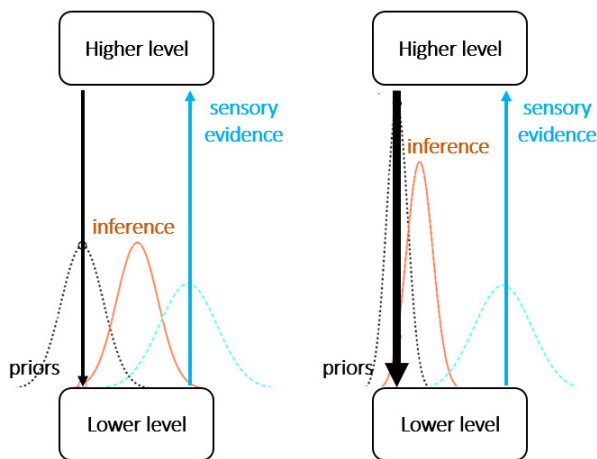
Figure 4. Overweighting of perceptual priors
Left: Perceptual inference integrates top-down priors (black) and bottom-up sensory evidence (blue) into a percept (posterior inference, orange), according to their precisions (narrowness of the underlying Gaussian distributions and thickness of the corresponding arrows). Right: Overly precise priors (narrow black distribution and thick black arrow) drive the inference away from sensory evidence (orange distribution shifted towards left), causing percepts without external stimulations – i.e. hallucinations.

This view is supported by empirical data demonstrating that both clinical (Kot & Serper, 2002) and non-clinical (Alderson-Day et al., 2017) voice-hearers, as well as psychosis-prone individuals (Teufel et al., 2015) favor prior knowledge over sensory information during perceptual inference. In a recent seminal paper (Powers et al., 2017), formally described and identified neural underpinnings of excessive prior-weighting in voice-hearers, independently of their clinical status. Specifically, using a conditioning paradigm, they showed that expectations (priors) of target presence – built upon learned associations between the target and accompanying stimuli – drive reports of perceiving the target even in its absence. By demonstrating a substantial overlap between the neural circuit commonly involved in AVH and the circuit underlying prior-driven hallucinations, their neuroimaging results suggest that AVH could essentially constitute a type of a prior, as has similarly been proposed by the work of others (Alderson-Day et al., 2017; Cassidy et al., 2018; Teufel et al., 2015; Zarkali et al., 2019). Collectively, these data show that perceptual inference in hallucinating individuals is predominantly driven by prior beliefs. However, this work has only employed simple behavioral tasks and has not addressed the relationship of prior beliefs to the perception of voices specifically. Additionally, it remains unclear which kinds of priors need to be over-weighted in order to experience AVH, especially to account for the immense phenomenological variability present among AVH.

1.4 Thesis at a glance

The general introduction is followed by a collection of articles that I produced during my doctoral studies. They are split in two conceptual parts. **PART I** describes two studies aimed at unraveling behavioral (**Study 1**) and neural (**Study 2**) mechanisms underlying self-voice perception. **PART II** consists of three studies in which I used the aforementioned robotic procedure (Blanke et al., 2014) to experimentally induce alterations in healthy participants' self-voice perception that have been proposed to underlie AVH (Frith, 1992; Powers et al., 2017). In **Study 3**, I investigated alterations in self-other voice discrimination and loudness perception, which I further related to interoception – specifically to processing of breathing and heartbeat signals (**Study 4**). In **Study 5**, I

investigated the effects of robotically-induced psychosis-like states on self- and other-voice detection. Thesis document is followed by a general discussion and three annexed articles, to which I contributed during my doctoral studies, albeit outside of my main research topic. Personal contributions for each study are summarized in the following section.

1.5   Personal contributions

For all the articles included in the main part of my thesis (Studies 1 – 5), my contribution included study design, data collection and analyses, as well as writing of the manuscript.

PART I

Study 1: Following the inner voice: Approximating internal self-voice representation with bone conduction. **Orepic, P.**, Kannape, O. A., Faivre, N. & Blanke, O. *In preparation*

Study 2: *EEG spatiotemporal patterns underlying self-other voice discrimination*. **Orepic, P.***, Iannotti G. R.*, Alcoba-Banqueri, S., Garin, D. F. A., Schaller, K., Michel, C. M. & Blanke, O. *In preparation*

PART II

Study 3: *Sensorimotor conflicts induce somatic passivity and louden quiet voices in healthy listeners.* **Orepic, P.**, Rognini, G., Kannape, O. A., Faivre, N.* & Blanke, O.* *Submitted and available on BioRxiv*

Study 4: *Breathing affects self-other voice discrimination in a bodily state with increased otherness.* **Orepic, P.**, Park, H.D., Rognini, G., Kannape, O. A., Faivre, N.* & Blanke, O.* *Submitted and available on PsyArXiv*

Study 5: Robotically-mediated sensorimotor stimulation induces identity-specific auditory-verbal hallucinations in healthy individuals. **Orepic, P.**, Bernasconi, F., Faggella, M., Faivre, N & Blanke, O. *In preparation*

APPENDIX

Supplementary study 1: *Enhancing analgesic neuromodulation through personalized immersive virtual reality in patients with chronic leg pain.* Solcà, M., Krishna, V., Young, N., Geist, T., Herbelin, B., **Orepic, P.**, Mange, R., Rognini, G., Serino, A., Rezai, A. & Blanke, O. *Pain (accepted)*

Personal contribution: technical development, data collection.

Supplementary study 2: *What it feels like to move via an intracortical brain machine interface (BMI).* Serino, A.*, Bockbrader, M.*, Colachis, S., Solcà, M., **Orepic, P.**, Bertoni, T., Dunlap, C., Eipel, K., Ganzer, P., Annetta, N., Sharma, G., Friedenberg, D., Sederberg, P., Faivre, N., Rezai, A.* & Blanke, O.* *Submitted*

Personal contribution: technical development, data collection.

PART I: Self-voice perception

2.1 Study 1: Following the inner voice: Approximating internal self-voice representation with bone conduction

Pavo Orepic[1], Oliver A Kannape[1], Nathan Faivre[2], Olaf Blanke[1, 3]

[1] Laboratory of Cognitive Neuroscience, Center for Neuroprosthetics and Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Switzerland

[2] Laboratoire de Psychologie et Neurocognition (LPNC), CNRS UMR 5105, Université Grenoble Alpes, France

[3] Faculty of Medicine, University of Geneva, Geneva, Switzerland

Corresponding author

Olaf Blanke

Bertarelli Chair in Cognitive Neuroprosthetics, Center for Neuroprosthetics & Brain Mind Institute, School of Life Sciences, Campus Biotech, Swiss Federal Institute of Technology (EPFL), 1012 Geneva, Switzerland

E-mail: olaf.blanke@epfl.ch

Tel: +41 (0)21 693 69 21

Fax: +41 (0)21 693 69 22

2.1.1   Abstract

Compared to extensive work on self-face and self-body perception, surprisingly few studies have investigated self-voice perception. Such underrepresentation is a consequence of the inability to match experimental self-voice stimuli, traditionally presented through air conduction, to the internal self-voice representation. Namely, natural speaking inevitably involves bone conduction of our voice, which alters its acoustical properties and augments audition with vibrotactile excitation, rendering self-voice perception multimodal. Here, we designed a sensitive self-other voice discrimination (SOVD) task with multimodal stimuli presentation through a bone conduction headset. In four studies, we demonstrated that bone conduction improves SOVD and increases skin conductance response to self-voice, additionally exploring the effects of familiarity and acoustical similarity to the other voice as well as previous self-voice exposure on SOVD. With this work, we propose a novel method for auditory self-identification and shed new light on the phenomenology of the self by portraying self-voice as a fundamentally multimodal construct.

Keywords

2.1.2   Introduction

Recent work has highlighted the importance of multisensory processing in constructing a representation of our body and the self (Blanke & Metzinger, 2009; Blanke, Slater, & Serino, 2015; Gallagher, 2000; Jeannerod, 2003; Park & Blanke, 2019). A continuous integration of various exteroceptive (e.g. visual, auditory, and tactile) and interoceptive signals (e.g. visceral, heartbeat, and respiration) serves as a basis for maintaining a stable own-body perception. Accordingly, experimental perturbation of spatio-temporal properties of multisensory signals can lead to altered states of bodily self-consciousness (BSC), such as disturbed self-identification, self-location or first-person perspective (Adler, Herbelin, Similowski, & Blanke, 2014; Aspell et al., 2013; Lenggenhager, Tadi, Metzinger, & Blanke, 2007; Pfeiffer, Schmutz, & Blanke, 2014; Preston, Kuper-Smith, & Ehrsson, 2016). It has further been shown that stimuli associated to the self are processed as special, i.e. differently from other (highly) familiar stimuli (Northoff et al., 2006; Qin, Wang, & Northoff, 2020; Sui & Humphreys, 2017). The two stimuli most associated to the self, and person identification in general (Blank, Wieland, & von Kriegstein, 2014), are own face and voice. Compared to extensive work focusing on self-face representations (reviewed by (Uddin, Kaplan, Molnar-Szakacs, Zaidel, & Iacoboni, 2005)), with the aim of unveiling the particularities of the self as opposed to other familiar and unfamiliar entities, the number of studies investigating specificities of own voice perception is surprisingly scarce. However, acquiring a better understanding of self-voice perception is of utmost importance, as erroneous self-voice misattribution is thought to account for auditory-verbal hallucinations (Frith, 1992; Frith & Done, 1989), the most common hallucination in schizophrenia, associated with high degrees of distress in the affected population (Harkavy-Friedman et al., 2003).

Most studies that compared recognition of self- and other voices indicated an inability to recognize our voice as well as other voices. In the early studies, the recognition rates for self-voice stimuli were strikingly low – e.g. 55% (Olivos, 1967) or even 38% (Rousey & Holzman, 1967). Similarly, two studies (Douglas & Gibbins, 1983; Gur & Sackeim, 1979) reported slower response times for self-voice, compared to other-voice stimuli. These authors proposed that this discrepancy might result from a difference in previous exposure to self-voice compared to other voices. This was directly addressed by (Rousey & Holzman, 1967), who demonstrated an increase in accuracy rates for participants who heard their recorded voices on a frequent basis, such as radio announcers. Contrary to the first wave of self-voice studies, today there are ceiling effects present throughout self-voice recognition tasks (Candini et al., 2018; Candini et al., 2014; Hughes & Nicholson, 2010; Rosa, Lassonde, Pinard, Keenan, & Belin, 2008). This shows that half a century after the first self-voice studies have been conducted, people learned to associate own voice recordings with themselves, arguably due to a higher exposure to own voice through contemporary technology (e.g. voice messages and video recordings). However, the differences in reported accuracy rates between self and other voices still persist in contemporary studies (Candini et al., 2018; Candini et al., 2014; Hughes & Nicholson, 2010), even in paradigms with distorted vocal stimuli (Allen et al., 2005, 2004; Rosa et al., 2008; Shuster, 1998). Despite being more exposed to self-voice recordings, they still do not sound natural to us and as a consequence, hearing our voice in a recording is often followed by feelings of unpleasantness and discomfort (Holzman & Rousey, 1966; Holzman, Rousey, & Snyder, 1966).

A decrease in ability to recognize our voice in self-voice recordings arguably arises due to a lack of bone conduction, which is inevitably present while speaking (Békésy, 1949; Reinfeldt, Östli, Håkansson, & Stenfelt, 2010). Due to the different propagation path-

way, bone conduction transforms the sound of our voice and, although extensively investigated (Dobrev et al., 2017; Pörschmann, 2000; Stenfelt, 2011; Stenfelt & Goode, 2005), the exact transfer function of bone-conducted compared to the air-conducted sound of our voice is still unknown (Stenfelt, 2016). However, it is assumed to represent a low-pass filter (Tonndorf, 1976; Wheatstone, 1827) and thus when we speak, compared to others, we hear our voice as lower. This was investigated by (Shuster & Durrant, 2003), who showed a preference for low-pass filtered, compared to unfiltered self-voice recordings while comparing real-time speech to delayed-recorded speech samples. However, other studies indicated preferences for different types of filters (Vurma, 2014; Won, Berger, & Slaney, 2014) and more systematic approaches could not converge to a single filter best approximating internal self-voice representation (Kimura & Yotsumoto, 2018; Maurer & Landis, 1990). Importantly, as physical transformation related to bone conduction is not present in self-voice recordings, the self-voice stimuli consistently present throughout self-voice studies do not correspond to the natural sound of our voice; hence, the findings of the studies demonstrating self-voice particularities compared to other voices should be taken with caution.

In addition to the physical transformation of the sound of our voice, during speech our voice is also inevitably accompanied by a vibrotactile excitation (Stenfelt, 2011) caused by bone conduction of the sound through the skull. Namely, when we speak, not only auditory, but also somatosensory (Ito, Tiede, & Ostry, 2009; Tremblay, Shiller, & Ostry, 2003) and vestibular (Emami et al., 2012; Todd, Cody, & Banks, 2000) processing can be involved. It is therefore possible that in order to increase familiarity to self-voice recordings, not only the physical properties of the sound of our voice should be approximated to the experience of speaking, but also the stimulus presentation involving somatosensory and vestibular excitation. A multimodal presentation of own voice can be achieved through commercial bone conduction headsets that have recently become commercially available and one study has even demonstrated that they enable better speech recognition compared to regular air-conduction headsets (Manning, Mermagen, & Scharine, 2017). Combining voice-morphing procedures (Kawahara, Morise, Banno, & Skuk, 2013), psychophysics, and auditory feedback delivered via bone conduction, we here propose a new experimental paradigm to study self-voice perception and contrast it with perception of familiar and unfamiliar voices. Building up on the multisensory accounts for bodily self-consciousness (Blanke & Metzinger, 2009; Blanke et al., 2015; Gallagher, 2000; Jeannerod, 2003; Park & Blanke, 2019), we tested whether multimodal presentation of own voice through bone conduction better approximates internal self-voice representation, compared to conventional air conduction of self-voices. Moreover, we designed an implicit measure of the association between bone conduction and self-voice by recording skin conductance response (SCR) to air- and to bone-conducted self- and other vocal stimuli. Namely, hearing own voice is known to elicit a higher SCR compared to hearing other voices, regardless of whether the voice is recognized as belonging to oneself (Douglas & Gibbins, 1983; Gur & Sackeim, 1979; Holzman et al., 1966; Olivos, 1967), suggesting that a SCR increase might serve an implicit measure of self-voice recognition. Thus, we hypothesized that with bone conduction the self-voice-related SCR increase would be higher compared to the air-conduction increase, indicating a preference of the bone-conducted sound of own voice even by the autonomic system.

In a series of four studies in independent cohorts, we further addressed other open questions related to self-voice phenomenology, ranging from low-level acoustical features to higher-level cognitive processes. First, it is unknown whether acoustical parameters that have been shown to account for discriminability of other voices (Baumann & Belin, 2010) also account for self-other voice discrimination (SOVD). Baumann and Belin identified a two-dimensional perceptual voice space (Baumann & Belin, 2010) in which similarly sounding voices are located close to each other and distances in such a voice-space were correlated with subjective distinctiveness of voices and BOLD responses in temporal voice areas (Latinus, McAleer, Bestelmeyer, & Belin, 2013). Here, we placed our participants' voices in an equivalent voice space and correlated self-other voice distances to SOVD task performance. Second, it is possible that, rather than comparing acoustical differences between self-voice and other voices, SOVD relies on matching a perceived voice to an internal self-voice representation, favoring familiarity mechanisms over acoustic processing. Here, we investigated the differences in SOVD dependent on other-voice familiarity and further explored whether self-voice is more confused with a familiar compared to an unfamiliar voice. Third, previous studies only speculated about the role of previous exposure to self-voice recognition. Here, we directly compared SOVD with and without immediate previous exposure to self-voice stimuli. Finally, it is unknown whether own voice is simply a (highly) familiar voice, or are there any perceptual characteristics specific to self-voice. Although some have identified differences between the processing of self-voice compared to familiar voices (Graux, Gomot, Roux, Bonnet-Brilhault, & Bruneau, 2015; Nakamura et al., 2001), to date, no study has explicitly explored differences between SOVD and familiar-other voice discrimination (FOVD). Here, we performed all the aforementioned self-voice assessments (e.g. voice-space analysis or previous exposure) also to familiar-voice, thereby determining the degree of the specificity of the effects for the self. All the assessments were also conducted both with air and bone conduction.

2.1.3 Methods

We investigated self-voice perception in a series of four studies in independent cohorts. Study 1 investigated differences in SOVD as a function of sound conduction (air, bone) and previous exposure to self-voice; Study2 extended this to FOVD. In Study 3, we set out to replicate studies 1 and 2 within a single, correctly powered cohort, as determined by power analysis of Study 1. Study 3 additionally contained self-familiar voice discrimination and a control self-voice recognition task, which involved no voice morphing. Study 4 investigated the effects of self-voice on SCR. Sample size for Study 4 was determined to match the number of experimental conditions.

2.1.3.1 Participants

Studies 1 and 2 each involved 16 participants. In Study 1, seven participants were male (mean age ± SD: 29.7 ± 5.5 years old) whereas 8 were male in Study 2 (28.5 ± 5.5 years old). For Study 3, participants were accompanied by an acquaintance (a friend) of the same gender and similar age, who also participated in the study, and it involved 52 participants (20 male, 26.5 ± 4.6 years old). In Study 4, 24 participants were tested (27.4 ± 3.5 years old), 14 of which were male. All participants were right-handed, reported no hearing deficits, and no history of psychiatric or neurological disorders. They were chosen from the general population and were naïve to the purpose of the study. Participants gave informed consent in accordance with institutional guidelines (protocol 2015-00092, approved by the Comité Cantonal d'Ethique de la Recherche of Geneva) and the Declaration of Helsinki, and received monetary compensation (CHF 20/h).

2.1.3.2 Procedure

2.1.3.2.1 Studies 1-3: Voice discrimination

Prior to participating in the studies, participants' voices were recorded while vocalizing the phoneme /a/ for approximately 1 to 2 seconds (Zoom H6 Handy recorder). Each recording was normalized for average intensity (-12 dBFS) and duration (500 milliseconds) and cleaned from background noise (Audacity software). Such preprocessed voice recordings were used to generate voice morphs spanning a voice identity continuum between two voices by using TANDEM-STRAIGHT (Kawahara et al., 2013) (e.g. a voice morph can be generated such that it contains 40% of person A's, 60% of person B's voice).

In Study 1, we morphed each participant's voice with the voice of a gender-matched unfamiliar person and participants performed the SOVD task using the resulting voice-morphs. For each voice morph, participants were instructed to indicate whether the voice they heard more closely resembled their own or someone else's voice by clicking on one of two buttons (2AFC). Six voice ratios (% self-voice: 15, 30, 45, 55, 70, 85) were chosen based on extensive pilot testing and were repeated 10 times within a block in a randomized order (total of 60 trials). Inter-trial intervals jittered between 1 and 1.5 seconds to avoid predictability of stimulus onset. Voice morphs were presented to participants either through bone-conducting headphones (Aftershokz Sports Titanium) or through laptop loudspeakers (air conduction). The study contained 4 experimental blocks, which differed based on the sound conduction type (air, bone) and whether there was immediate previous exposure to self-voice, i.e. whether the unmorphed self-voice was presented to participants prior to the block (yes, no). In the first two blocks, participants performed the task without having previously heard the recording of their voice, once with each type of sound conduction. This way, we ensured that the task performance was based on their internal self-voice representation and unaffected by any bias introduced by hearing a recording of one's own voice. The order of air- and bone-conduction blocks was counterbalanced across participants and for both parts of the experiment (with and without previous exposure to self-voice). Study 2 followed the same procedure, except that a familiar voice was morphed with the unfamiliar voice, instead of the self-voice. Thus participants performed a familiar-unfamiliar voice discrimination task. The familiar voice belonged to a male person to whom all participants were acquainted, the unfamiliar voice was the voice used in Study 1.

Study 3 consisted of two auditory tasks – a voice discrimination and a self-recognition task. Both tasks contained three types of voice stimuli – the participant's voice (self), his/her acquaintance's voice (familiar other) and the voice of the unknown third person (unfamiliar other). During both tasks, participants could hear the voice stimuli either through bone-conducting (Aftershokz Sports Titanium) or regular, air-conducting headphones (Bose QC20). Both were installed on participants' heads before the beginning of the experiment and matched for loudness at lower sound intensities, such that vibrational tactile sensations resulting from bone conduction could not be perceived, resulting in participants being unable to determine the source of the auditory stimuli throughout the experiment. This represents a methodological improvement compared to Study 1, where air conduction employed loudspeakers instead of headphones.

The self-discrimination task contained three types of experimental blocks: i) self-unfamiliar, ii) familiar-unfamiliar, and iii) self-familiar – which were identical to studies 1 and 2. The target voice (self- or familiar-voice) was not shown to participants prior to the block, and each of the three blocks (self-unfamiliar, familiar-unfamiliar, self-familiar) was conducted once for each type of sound conduction (air, bone). The order of the four blocks containing morphs with an unfamiliar voice (self-unfamiliar and familiar-unfamiliar, both with air and bone conduction) was pseudorandomized across participants, whereas the remaining two self-familiar blocks were counterbalanced across participants and always conducted at the end. This was to balance the exposure to self- and to familiar-voice for their discrimination from the unfamiliar voice.

In the self-recognition task, unbeknown to participants, the stimuli consisted only of unmorphed voices (self, familiar and unfamiliar) and in each trial, participants were instructed to indicate whether the voice they hear sounded like their own by pressing a button. Similar to the self-discrimination task, there were two experimental blocks, one for each form of sound conduction (air, bone), counterbalanced across participants. Each voice type was randomly repeated 10 times within the block and intertrial interval jittered between 1 and 1.5 seconds. Voice recognition task served as control to identify whether participants were able to recognize their recorded voice even without morphing. Thus, it was always performed at the end of experiment, not to affect the performance in the discrimination task by previous exposure with unmorphed voice recordings, i.e. to ensure that participants performed self-other discrimination relying on the internal representation of their own voice. Figure 5 Illustrates the design of all 3 studies. All studies were performed in MATLAB 2017b with Psychtoolbox library (Kleiner et al., 2007).
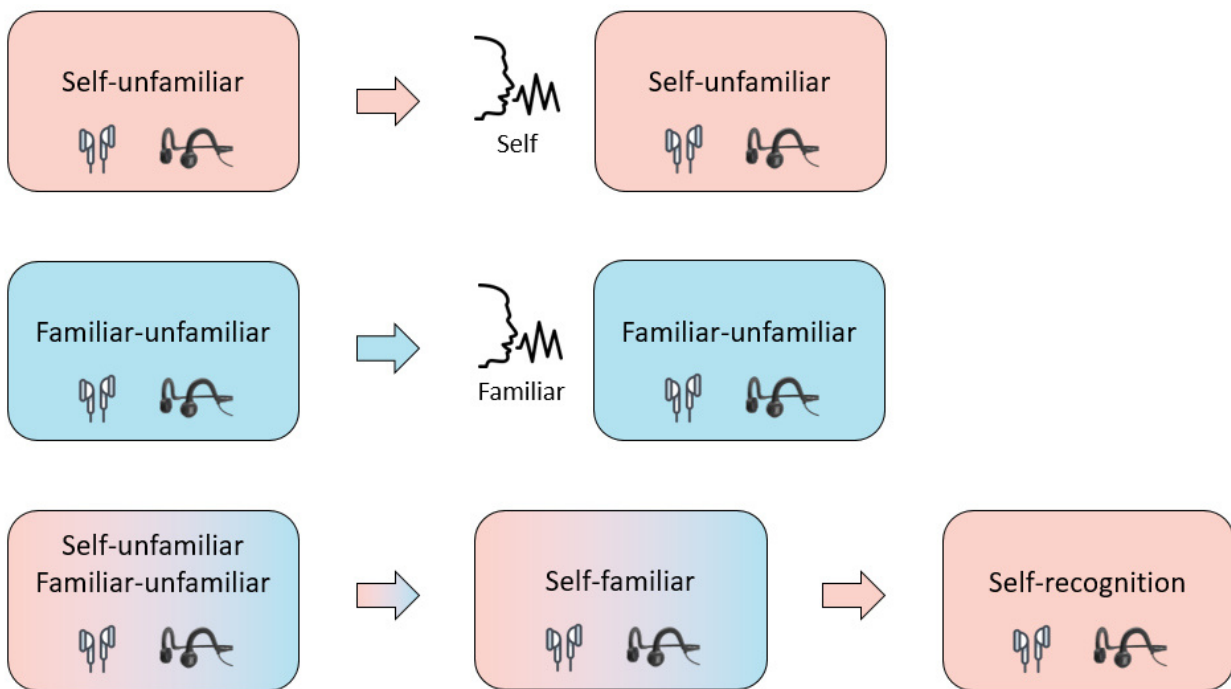
Figure 5. Experimental design.
Top row illustrates the design of Study 1, middle of Study 2 and bottom of Study 3. Blocks represent different types of auditory tasks. In all studies, all tasks were performed both with air and with bone conduction (white and black icons, respectively). In the first two studies, voice discrimination task was first performed without and then with having the target voice shown prior to the task. In Study 3, the target voice was never shown. The self- and familiar-voice were first discriminated against the unfamiliar voice and then against each other. Self-voice recognition task over the unmorphed voices was conducted at the end of Study 3.

2.1.3.2.2   Study 4: Skin conductance response
In Study 4, participants performed a voice detection task. They were exposed to short bursts of pink noise and were instructed to report whether they heard a voice in the noise (by clicking on a button after the noise stimulus ended). In 71.5% of trials, there was a voice presented with a higher sound intensity than the noise, ensuring that it could always be clearly heard by participants. There were four experimental blocks that differed based on sound conduction (air, bone) and voice type (self, other), and that were counterbalanced across participants, resulting in each participant (N = 24) having a different order of experimental conditions. Prior to the experiment, participants' voices were recorded while saying nine short words in French (supplementary material). The other voice was unfamiliar to participants. Voice recordings were standardized for duration (500 milliseconds) and loudness (-12 dBFS).

Each word was presented five times within a block together with 18 catch-trials containing no voice within a noise, both in a randomized order, resulting in 63 trials per block. Following participants' response in each trial (a button click), an inter-trial interval jittered between 1 and 1.5 seconds. Each burst of noise lasted for 3.5 seconds and voice onset randomly occurred in a period between 0.5 and 2.5 seconds after the noise onset, ensuring a minimum of 0.5 seconds of noise before and after the presentation of voice recording. Together, this design resulted that each trial contained on average 1.5 seconds before voice onset (min-max values: 0.5 – 2.5 seconds) and 3.5 seconds after voice onset (1.5 - 5.5 seconds).

Electrodermal activity (EDA) was collected continuously at a sampling rate of 2000 Hz throughout the experiment (Biopac MP36R system) from middle phalanges' palmar surfaces of ring and little fingers of the left hand. The signal was bandpass-filtered between 0.01 and 35 Hz and then down-sampled to 80 Hz. For each trial, a trigger was recorded for trial onset and for trial offset, as well as for voice onset. The filtered signal was extracted for the entire duration of each trial with the aim of investigating SCR to voice onset.

2.1.3.3   Statistical analysis

Performance in studies 1 and 2 task was analyzed with mixed-effects binomial regressions with Response as dependent variable, two fixed effects with an interaction term – Conduction (air, bone) and Previous Exposure (yes, no) – and a fixed effect of Voice Morph (15, 30, 45, 55, 70, 85). The Response-variable indicated whether participants identified the presented voice as their own (Study 1) or as their acquaintance's voice (Study 2).

In Study 3, the performance in the self-recognition task was analyzed with mixed-effects binomial regressions with Response as dependent variable and two fixed effects with an interaction term: Conduction (air, bone) and Voice (self, familiar, unfamiliar). The Response-variable indicated whether participants identified the presented voice as their own. The effect of sound conduction on each type of voice discrimination (self-unfamiliar, familiar-unfamiliar, self-familiar) was analyzed with mixed-effects binomial regressions with Response as dependent variable and Conduction (air, bone) and Voice Morph (15, 30, 45, 55, 70, 85), together with a two-way interaction, as fixed effects. The Response-variable indicated whether participants perceived a voice morph as sounding more like their own (self-unfamiliar, self-familiar) or their acquaintance's voice (familiar-unfamiliar). For all mixed-effects regressions in all studies, random effects included a by-participant random intercept and by-participant random slopes for the main effects were added following model selection based on maximum likelihood. Trials with reaction times greater or smaller than two interquartile ranges from the median for each participant were considered as outliers and excluded. Participants who could not recognize their voice in more than half of the self-voice trials in the self-recognition task (where there was no voice morphing), were considered as outliers and excluded from the self-discrimination task analysis. Additionally, a linear mixed-effects regression with Reaction Times as a dependent variable and the same fixed and random effects was performed for all studies, with the polynomial expansion of the Voice Morph variable to level 2 (supplementary material).

For the self-recognition task, we additionally explored whether self-voice was more confused with the familiar or with the unfamiliar voice. For that purpose, we correlated the inverse accuracy for the self-voice trials (indicating the rate to which self-voice was not identified as own voice, i.e. miss rate) with the rate to which familiar- and unfamiliar-voice were misperceived as self-voice (i.e. false alarm rate). Pearson and Filon's z-test for comparing two correlations based on dependent groups with overlapping variables(Pearson & Filon, 1898) was used to compare these two correlations (miss rate with two types of false alarm rates – familiar-as-self and unfamiliar-as-self misperception). The two misperceptions were also correlated with each other. Where significant, separate correlations were then conducted for and compared between the two forms of sound conduction (air, bone).

To investigate whether SOVD supports prototype-based coding of voices(Baumann & Belin, 2010), we correlated the performances in SOVD tasks (both self-unfamiliar and self-familiar) with the self-other distances in the two-dimensional perceptual voice space identified by (Baumann & Belin, 2010). As the perceptual voice-space differs for male and female voices (Baumann & Belin, 2010), we placed our voice recordings in the corresponding gender-based space and correlated voice distances with task performance separately for each gender. Thus, for each voice recording, we extracted the fundamental frequency (F0) and five formants (F1-F5) using Praat software (Boersma & Weenink, 2020) and computed its voice-space coordinates, corresponding to source (x coordinate) and filter (y) components of voice production (males: x = log(F0), y = log(F5 – F4); females: x = log(F0), y = log(F1)). The coordinates were first transformed into z-scores, after which the voice spaces were normalized for the other voice, such that other-voice coordinates were subtracted from self-voice coordinates in each self-other voice pair. This resulted in a coordinate system where Euclidean distance to the origin represented self-other voice distance in z-score units. Distances to the origin (self-other perceptual voice distances) were then correlated with the percentage of correct responses in self-other task. In the same way, we created familiar-other voice space and compared familiar-other distances with familiar-other task performances. Finally, we ap-

plied Pearson and Filon's z-test for comparing two correlations to assess differences in relationship between voice-space distance and task performance for self-other and familiar-other distinctions, separately for each voice-space coordinate (source, filter). Significant correlations were ran again for and compared between the two forms of sound conduction (air, bone).

In Study 4, per-trial SCR was normalized by subtracting the mean value occurring before the voice onset. Subsequently, the maximal value occurring after the voice onset was extracted for each trial and served as a dependent variable in a linear mixed-effects regression analysis. Fixed effects in the regression were Voice (self, other), Conduction (air, bone) and Trial (1-63), all related with an interaction term. The trial number was added as a covariate as the data displayed a steady decrease in absolute EDA values in most participants with SCR mainly varying at the beginning of experimental blocks.

Statistical tests were performed with R (R Core Team, 2020), using the lme4 (Bates, Mächler, Bolker, & Walker, 2015), lmerTest (Kuznetsova, Brockhoff, & Christensen, 2018), and cocor (Diedenhofen & Musch, 2015) packages. The results were illustrated using sjplot (Lüdecke, 2018) and ggplot2 (Wickham, 2016) packages.

2.1.4 Results

2.1.4.1 Study 1: Self-unfamiliar discrimination

A mixed-effects binomial regression in the self-unfamiliar discrimination task with Response as a dependent variable revealed main effects of Conduction (estimate=-0.47, Z=-2.96, p=.003), Previous Exposure (estimate=-0.5, Z=-4.64, p<.001) and Voice Morph (estimate=0.55, Z=22.67, p<.001), indicating more 'self' responses for bone compared to air conduction, as well as for the blocks without compared to the blocks with previous exposure. The ratio of 'self' response increased with increased amount of self-voice present in voice morphs. Moreover, the analysis yielded a significant interaction between Conduction and Previous Exposure (estimate=0.43, Z=2.85, p=.004). In order to investigate the nature of the interaction, we ran a separate mixed-effects binomial regression for each type of Previous Exposure. For the blocks with previous exposure to self-voice, only the main effect of Voice Morph was significant (estimate=0.9, Z=7.28, p<.001). There was no main effect of Conduction (estimate=0.4, Z=1.4, p=.16), nor did the Conduction interact with the effect of Voice Morph (estimate=-0.12, Z=-1.63, p=.11). On the contrary, besides the main effect of Voice Morph (estimate=0.59, Z=4.39, p<.001), the analysis for the blocks without previous exposure to self-voice also revealed a significant interaction between Voice Morph and Conduction (estimate=-0.25, Z=-3.46, p<.001), indicating a steeper slope for the psychometric curve fitted for bone conduction, compared to the curve fitted for air conduction. The main effect of Conduction was not significant (estimate=0.39, Z=1.38, p=.17). Collectively, the results of Study 1 indicate that participants perform better in the self-other discrimination task (1) with previous exposure to own voice, (2) with voice morphs presented through bone conduction and (3) that bone conduction improves SOVD more without previous exposure to own voice (Figure 6, top).

2.1.4.2 Study 2: Familiar-unfamiliar discrimination

A mixed-effects binomial regression in the familiar-unfamiliar discrimination revealed a main effect of Voice Morph (estimate=1.16, Z=30.53, p<.001). The effects of Conduction (estimate=-0.03, Z=-0.22, p=.826) and Voice Presentation (estimate=-0.54, Z=-1.17, p=.242) were not significant, nor was the interaction between them (estimate=0.27, Z=1.41, p=.159). Familiar-unfamiliar discrimination was unaffected by sound conduction and previous exposure to the familiar voice (Figure 6, bottom).

2.1.4.3 Study 3: Self-recognition and self-confusion

A mixed-effects binomial regression in the self-recognition task showed higher accuracy for self-voice trials (i.e. hit rate, mean = 0.79, 95% CI = [0.72, 0.85]) both compared to familiar-voice (i.e. familiar false alarm rate, 0.17, [0.11, 0.23]; estimate=-3.23, Z=-17.48, p<.001) and unfamiliar-voice trials (i.e. unfamiliar false alarm rate, 0.13, [0.08, 0.18]; estimate=-3.44, Z=-17.93, p<.001). There were no differences in false alarm (FA) rates between familiar and unfamiliar voices (estimate=-0.21, Z=-1.1, p=.273) (Figure 7, left). The main effect of Conduction was not significant (estimate=0, Z=0.03, p=.98) nor was there a Conduction by Voice interaction (estimate=-0.13, Z=-0.46, p=.643).

Pearson's product-moment correlation did not show a significant relationship between the two false alarm rates (r=-0.07, 95% CI = [-0.34, 0.21]; t(50)=-0.49, p=.624), showing that participants either misperceived familiar or unfamiliar voice as their own, and that other-to-self voice misperceptions were not related (Figure 7, middle). However, correlation analysis identified a significant positive relationship (r=0.67, 95% CI = [0.48, 0.79]) between miss rate and familiar-FA rate (t(50)=6.31, p<.001) while there was no

significant relationship (t(50)=1.46, p=.151) between miss rate and unfamiliar-FA rate (r=0.2, 95% CI = [-0.07, 0.45]) (Figure 7, right). Pearson and Filon's z-test identified a stronger relationship between miss and familiar-FA compared to unfamiliar-FA rates (z=2.86, p=.004), indicating that participants were confusing own voice more to the familiar compared to the unfamiliar voice. Neither familiar-FA nor unfamiliar-FA rates were affected by sound conduction (supplementary material).
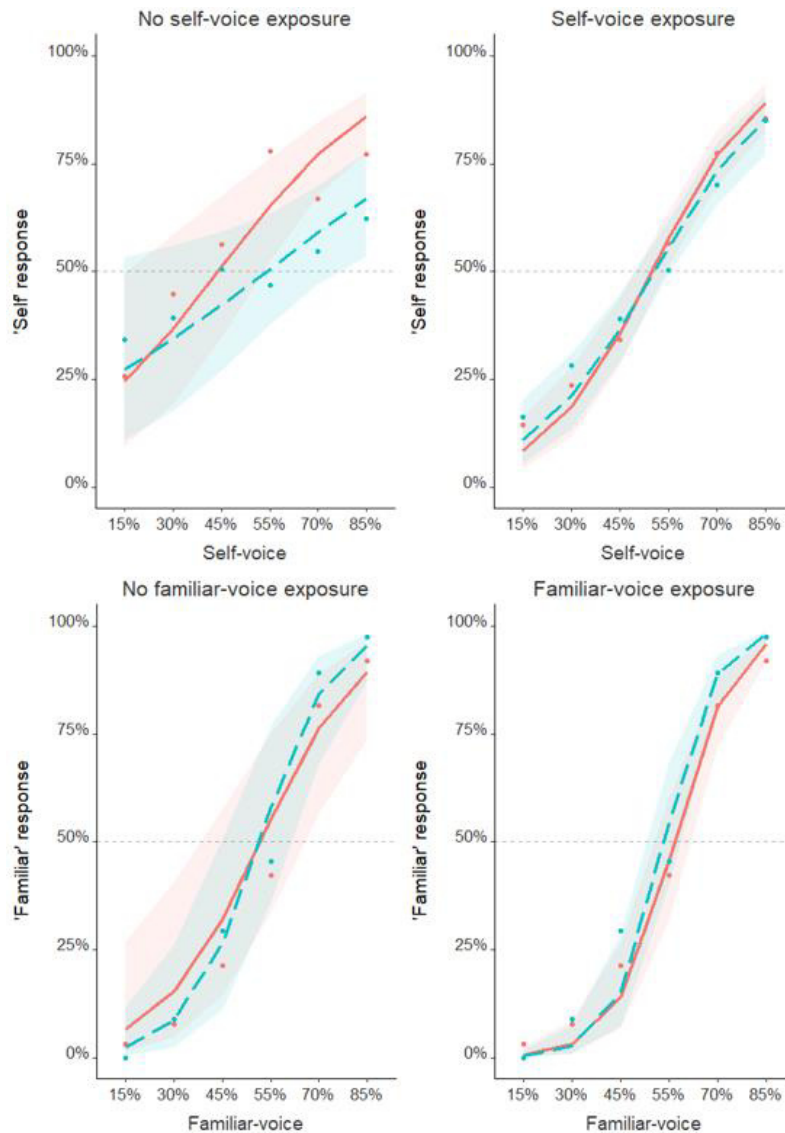
Figure 6. Studies 1 and 2.
Psychometric curves fitted for two forms of sound conduction (air, bone) during studies 1 (top) and 2 (bottom). The abscissa indicates the percentage of the self/familiar voice present in a voice morph and the ordinate indicates the rate at which the corresponding voice morph was perceived as more resembling the self/familiar voice. The shaded areas around each curve represent the 95% confidence intervals. Left plots indicate perception for the blocks without and right plots for the blocks with previous exposure to the target voice prior to the task. Bone conduction improved self-unfamiliar discrimination only when participants were not previously exposed to their voice before the task (top). No such effects were observed for familiar-unfamiliar discrimination (bottom).
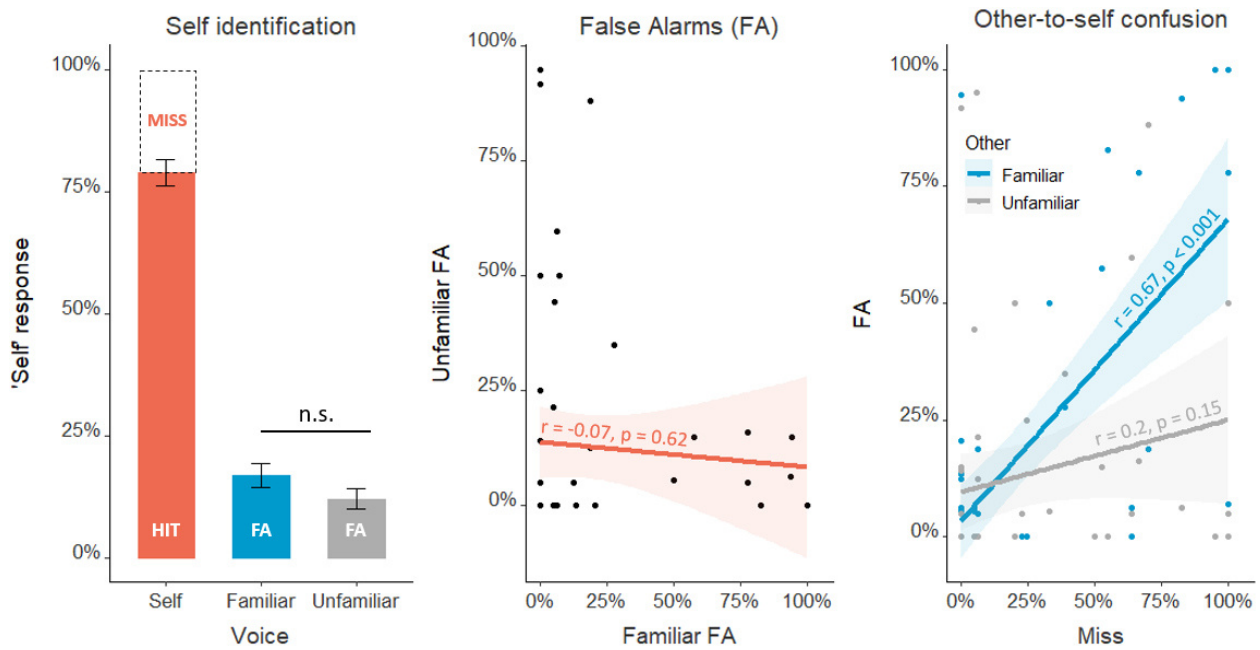
Figure 8. Study 3: Self-confusion.

The bar plot (left) indicates mean self-identification rates occurring for each type of voice stimuli – hit rate for self and false alarm (FA) rates for familiar and unfamiliar voices – whereas the regression plots indicate relationships between false alarm rates for familiar and unfamiliar voice with each other (middle), and with the miss rate for self-voice (right). Bar-plot whiskers and shaded areas around linear regressions indicate 95% confidence intervals. Although the absolute rate at which familiar and unfamiliar voices were misperceived as self-voice did not differ (left) and were not related to each other (middle), only the familiar voice misperception was related to a decrease in self-recognition (right).
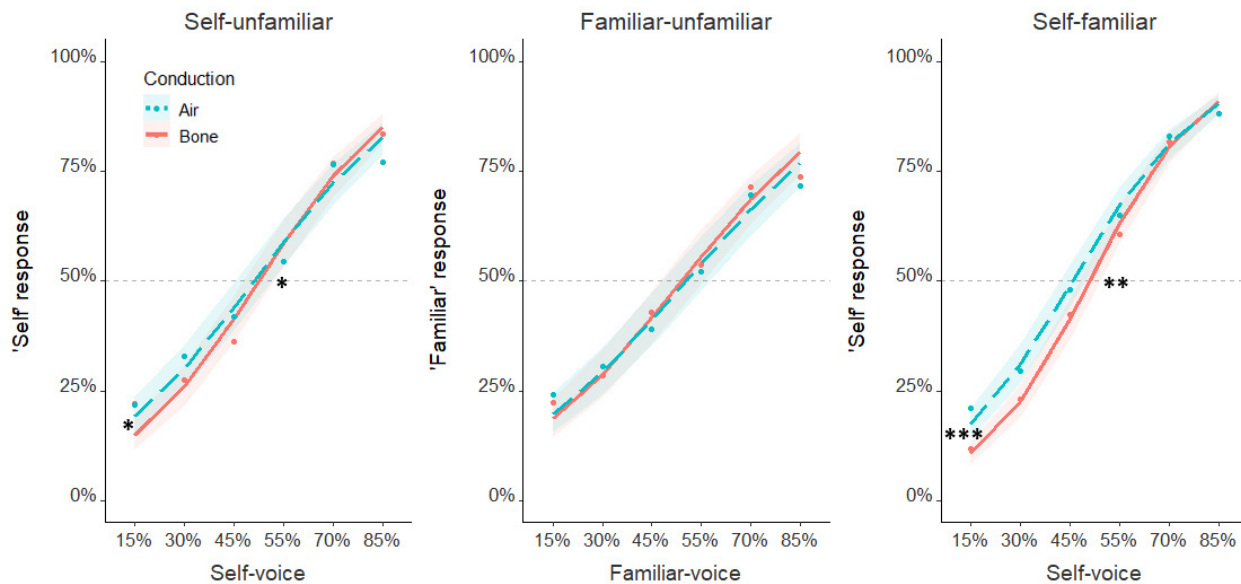
Figure 7. Study 3: Voice discrimination.

Psychometric curves fitted for two forms of sound conduction (air, bone) during all three voice-discrimination tasks (left: Self-unfamiliar; middle: Familiar-unfamiliar; right: Self-familiar). The abscissa indicates the percentage of the self/familiar voice present in a voice morph and the ordinate indicates the rate at which the corresponding voice morph was perceived as more resembling the self/familiar voice. The shaded areas around each curve represent the 95% confidence intervals. Asterisks in the lower end of the curves indicate a significant difference in intercepts, whereas asterisks in the middle of the curves indicate a significant difference in slopes. Intercepts were lower and the slope was steeper for the curves fitted for bone conduction, but only in the self-related tasks. *: p<0.05, **: p<0.01, ***: p<0.001

2.1.4.4 Study 3: Self-other discrimination

Nine participants who were unable to recognize their own unmorphed voice in more than a half of the self-voice trials during self-recognition task were excluded from the voice discrimination task analysis. A mixed-effects binomial regression in the self-unfamiliar discrimination revealed a main effect of Conduction (estimate=0.4, Z=2.44, p=.015) – showing higher intercepts for air conduction – and a main effect of Voice Morph (estimate=0.7, Z=22.13, p<.001). A two-way interaction between the effects of Conduction and Voice Morph was also significant (estimate=-0.1, Z=-2.26, p=.024), indicating a steeper curve and thus a better performance for bone conduction compared to air conduction (Figure 8, left). Conduction did not affect performance in familiar-unfamiliar discrimination (estimate=0.12, Z=0.75, p=.452), nor did it significantly interact with the effect of Morph (estimate=-0.05, Z=-1.12, p=.263). Only the main effect of Morph was significant for the familiar-unfamiliar discrimination (estimate=0.56, Z=19.09, p<.001) (Figure 8, middle). Analysis for the self-familiar discrimination showed the same, but more significant effects observed in self-unfamiliar discrimination – i.e. both a main effect of Conduction (estimate=0.7, Z=3.97, p<.001) and Morph (estimate=0.89, Z=25.15, p<.001), as well as their interaction (estimate=-0.13, Z=-2.71, p=.008) (Figure 8, right). Bone conduction improved the performance in the voice discrimination task only if the task involved self-voice morphs.

2.1.4.5 Study 3: Prototype-based coding of self-voices

Participants' voices were placed in perceptual voice spaces as defined by (Baumann & Belin, 2010) (Figure 9, left). Correlation analysis indicated a positive association between self-other voice distances and self-other task performance only for male participants (males: r=0.41, 95% CI = [0.11, 0.64]; t(38)=2.76, p=.009; females: r=-0.02, 95% CI = [-0.27, 0.23]; t(58)=-0.17, p=.869) (Figure 9, right). Neither sound conduction (air, bone) nor the type of other voice (familiar, unfamiliar) affected the relationship between task
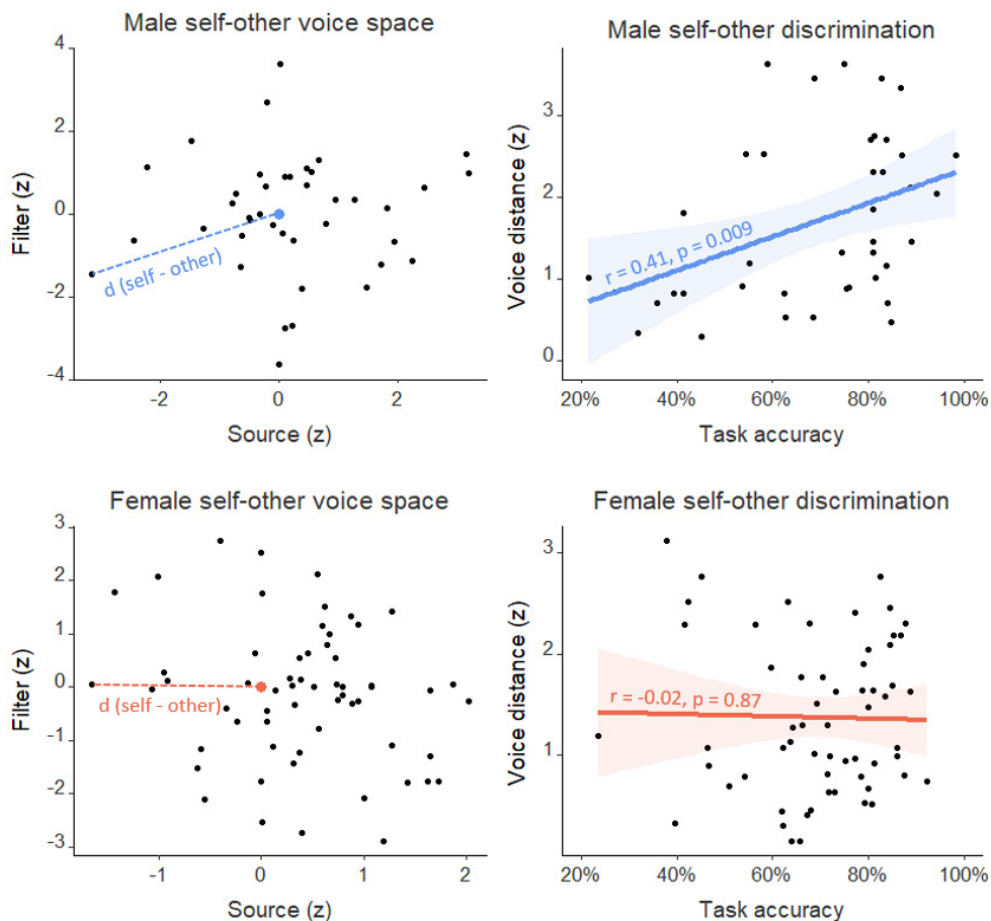
Figure 9. Prototype-based coding for self-voices.
Upper plots refer to male and lower plots female participants. Left plots indicate perceptual voice spaces where the origin (enlarged dot) represents the other voice in self-other discrimination tasks. Distance to the origin (dashed line) thus represents each participant's self-other voice distance in z-score units. On the right plots, self-other distances were correlated to the self-other task performances. Shaded areas around linear regression indicate 95% confidence intervals. Positive relationship between perceptual voice distance and task performance was found only for male participants.

performance and self-other distance (supplementary material).

We further correlated task performance of male participants separately with source and filter distances, and identified differences in reliance on filter component between SOVD and FOVD. Specifically, Pearson and Filon's z-test revealed a significant difference in correlations between filter component distances and task performances for self-unfamiliar (r=0.35, p=.132) and familiar-unfamiliar (r=-0.35, p=.164) distinctions (z=2.41, p=.016) (Figure 10, middle). No such difference in correlations was found for source component distances (self-unfamiliar: r=0.33, p=.156; familiar-unfamiliar: r=0.39, p=.086; z=-0.22, p=.819) (Figure 10, left). Interestingly, the observed difference in filter-component correlations was more pronounced for bone conduction (z=2.89, p=.004) whereas filter-distances of self-unfamiliar and familiar-unfamiliar distinctions did not correlate differently when the task was performed through air conduction (z=1.64, p=.11) (Figure 10, right).
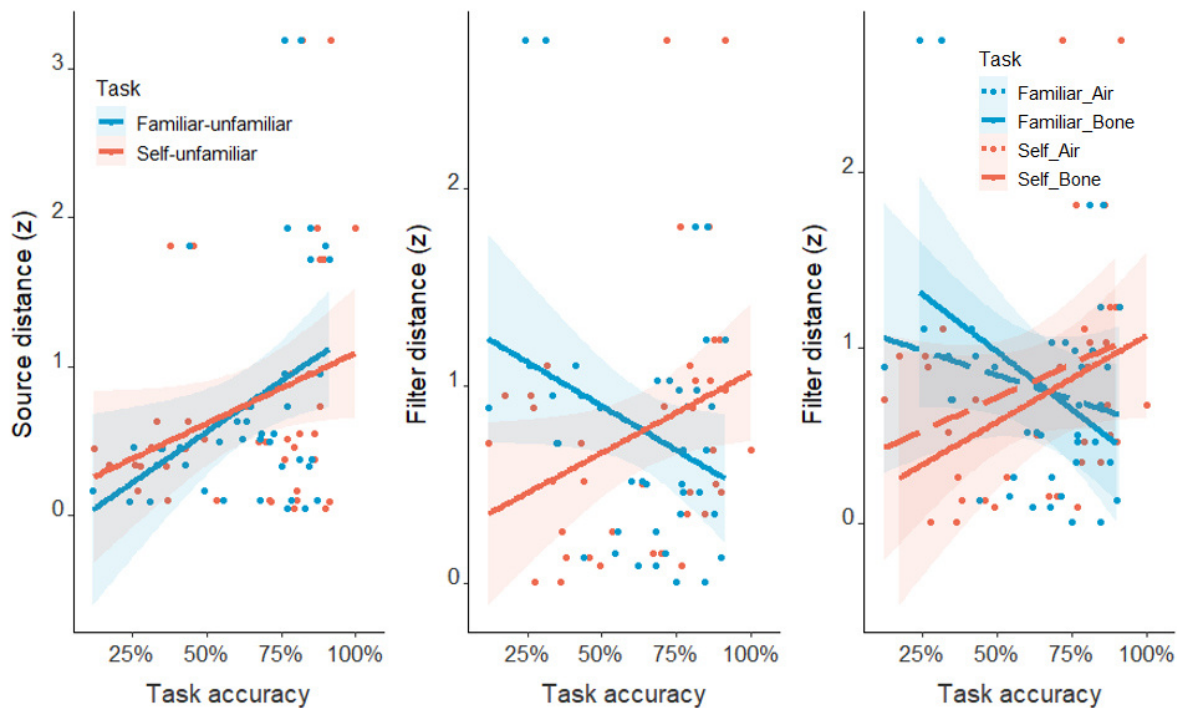
Figure 10. Contribution of filter and source components.
Different contributions of source and filter components of the source-filter perceptual voice space to self-unfamiliar and familiar-unfamiliar distinctions. Accuracy of the two distinctions was correlated with source (left) and filter (middle) component voice distances. The filter component was additionally compared both for bone and air conduction (right). Shaded areas represent 95% confidence intervals. The difference in relationships was significant only for the filter component, the effect enhanced for bone compared to air conduction.

2.1.4.6   Study 4: Skin conductance response (SCR)

Average SCR for all experimental conditions is illustrated at the top of Figure 11. A mixed-effects linear regression for max values of SCR revealed a main effect of Conduction (estimate=-0.058, df=4040, t=-3.53, p<.001) – showing a higher SCR for air compared to bone conduction. There was a main effect of Trial (estimate=-0.001, df=4041, t=-4.6, p<.001) – demonstrating a decrease of SCR with every following trial in the block – and it significantly interacted with the effect of Conduction (estimate=0.001, df=4040, t=2.24, p=.025) – indicating that this decrease was steeper with bone conduction. We did not observe a main effect of Voice (estimate=-0.027, df=4040, t=-1.64, p=.101) nor its interaction with the effect of Trial (estimate=0.001, df=4040, t=1.33, p=.183). Crucially, there was a significant interaction between effects of Conduction and Voice (estimate=0.053, df=4039, t=2.29, p=.022). Post-hoc investigation of the interaction indicated that SCR was higher for self-voice compared to other voice only with bone conduction (estimate=0.026, df=2016, t=2.03, p=.042), while with air conduction there were no significant differences between SCR for the two voices (estimate=-0.027, df=2003, t=-1.37, p=.169) (Figure 11, below). There was no significant three-way interaction between effects Conduction, Voice, and Trial (estimate=-0.001, df=4039, t=-1.68, p=.094).
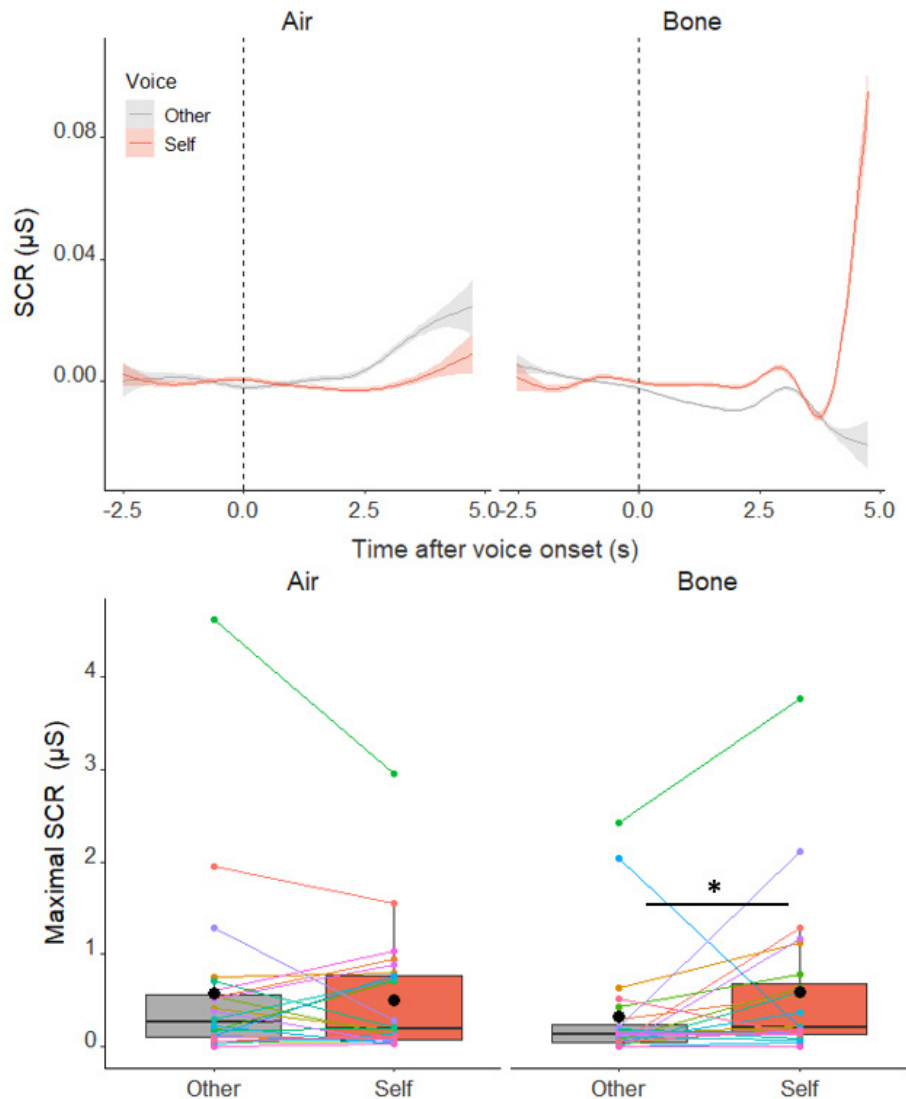
Figure 11. SCR response.

SCR response. Top: Average SCR to onset of self and other voices (vertical dashed line) presented through air (left) and bone conduction (right). Shaded areas around curves represent 95% confidence intervals. Bottom: Colored dots and lines represent each participant's maximal EDA value of the evoked response in the corresponding condition. Boxplots indicate interquartile ranges, horizontal lines inside boxplots median and big black dots mean values. SCR to self and to other voice differed only with bone conduction. *: p<0.05

2.1.5 Discussion

In three studies, we demonstrated that participants were better in discriminating their own from someone else's voice when they heard self-other voice morphs through bone as compared to air conduction. This advantage was more prominent in morphs that contained less self-voice features and when the task was performed without previous exposure to self-voice. No such effects were observed for discrimination of familiar from unfamiliar voices. We further showed that self-voice is more confused with a familiar than with an unfamiliar voice, regardless of acoustic similarity to those voices. Acoustical parameters accounting for discriminability of unfamiliar voices were, however, related to SOVD in male participants and, moreover, this relationship was more prominent for higher formants. Finally, in the fourth study, we observed a higher evoked response in skin conductance for hearing self-voice, compared to other voices, but only with bone conduction. Collectively, our findings demonstrate the importance of bone conduction with respect to self-voice perception and further dissociate processes underlying SOVD, specifically those relying on acoustical similarity and on familiarity of the other voice.

Our voice is an integral part of our self and the corporeal awareness of the self relies on multisensory integration of bodily signals (Blanke & Metzinger, 2009; Blanke et al., 2015; Gallagher, 2000; Jeannerod, 2003; Park & Blanke, 2019). Accordingly, self-voice perception is multimodal and besides auditory, it often involves somatosensory and vestibular afferences (Emami et al., 2012; Tremblay et al., 2003). However, this applies only to a natural scenario in which we hear our voice while we speak. Passively perceived self-voice that is traditionally presented through air conduction lacks such multimodal excitation and it also sounds differently to us, since it is not (low-pass) filtered as a result of passing through the skull. This leads to a reduced ability to recognize ourselves in air-conducted recordings (Allen et al., 2005; Gur & Sackeim, 1979; Hughes & Nicholson, 2010; Rosa et al., 2008; Schuerman, Meyer, & McQueen, 2015; Shuster, 1998), which is often associated with affective disturbances related to a discrepancy between the presented self-voice and its internal representation (Holzman & Rousey, 1966). Here, we showed that such discrepancy could be reduced when self-voice is presented through a commercial bone-conduction headset, by demonstrating and replicating an increase of performance in SOVD tasks with bone, compared to air conduction. Bone-conduction advantage was more prominent in self-familiar, compared to self-unfamiliar task. The reason for this might be simply that this task always occurred at the end of the experiment, whereas self-unfamiliar task blocks were counterbalanced with familiar-unfamiliar task, as our main goal was to contrast FOVD with SOVD for the same, unfamiliar other voice. Importantly, bone-conduction advantage did not extend to FOVD, divorcing bone-conduction advantage from the effects of voice familiarity. We argue that the reason for observing such self-specificity mainly lies in vibrotactile excitation accompanied by bone conduction, and only to a lesser extent in physical transformations to self-voice stimuli (e.g. deeper sound due to filter properties of bone tissues). This claim is mainly supported by a lack of differences in performance between bone and air conduction in FOVD tasks. Namely, if only physical transformations were to account for the bone-conduction effect in SOVD, we would have observed the opposite effect in FOVD – i.e. the advantage of air conduction for familiar voices, as we are not used to hearing familiar voices transformed in such a way. This suggests that self-voice is essentially a multimodal construct and consolidates the special nature of the self in perception of voices as it has been shown for perception of faces (Alzueta, Melcón, Poch, & Capilla, 2019). Moreover, as multisensory presentation seems to be the determining factor for improving SOVD, these findings might be of further relevance to unraveling the phenomenology auditory-verbal hallucinations (Frith, 1992; Frith & Done, 1989), as they have been proposed to arise as an impairment in SOVD.

With our investigations of previous exposure to self-voice, we further associated the multimodal properties of self-voice to the internal self-voice representation. Specifically, we narrowed the bone-conduction advantage down to SOVD tasks performed without previous exposure to unmorphed self-voice stimuli. Namely, by hearing an unmorphed recording of their voice prior to task execution, participants could have created an arbitrary strategy of recognizing that specific voice recording in a voice morph, regardless of whether they actually associated the recording with themselves. Thus, without previous exposure, there was no external reference and participants had to rely on their internal self-voice representation to complete the task. As bone conduction facilitated task performance only in this scenario (i.e. there were no differences in performance between bone and air conduction with previous self-voice exposure), we suggest that bone-conducted self-voice stimuli better corresponded to the internal self-voice representation. Crucially, bone conduction did not affect performance in blocks without previous exposure to familiar voice, suggesting that sound conduction is not related to internal familiar voice representation as it is to internal self-voice representation.

The association between bone conduction and internal self-voice representation was also observed from another, autonomic perspective, by detecting an increase in SCR for self-voice only when it was presented through bone conduction. Several research groups demonstrated that hearing own voice elicits a higher SCR compared to hearing other voices (Douglas & Gibbins, 1983; Gur & Sackeim, 1979; Holzman et al., 1966; Olivos, 1967), and similar was observed for seeing own face (Ameller et al., 2015; Sugiura et al., 2000). Interestingly, in these studies, SCR increased to self-related stimuli (both voice and face) even when they were erroneously misattributed to someone else, suggesting that SCR increase might constitute an implicit measure of self-recognition, occurring at the level of autonomic processing. Here we demonstrate that such an implicit measure of self-voice recognition could further be sensitized with bone conduction. Namely, although with our design most participants did not elicit an evoked SCR to voices (neither self or other), and although SCR occurred only in the beginning of experimental blocks (probably due to habituation and task monotony), bone conduction managed to facilitate a weak but significant self-related increase, suggesting its preferment over air conduction, even by the autonomic system.

We further investigated the contribution both of familiarity processing and acoustic differences to SOVD. On one hand, our self-confusion results show that self-voice perception inevitably involves some familiarity processing. Specifically, a failure to recognize own voice was correlated with familiar-to-self, and not with unfamiliar-to-self voice misattribution, regardless of acoustic similarity between the three voices. This suggests that in self-voice recognition, memory-related template-matching mechanisms might outweigh the processing of acoustical properties of the heard voice. On the other hand, our voice-space analysis indicates that, to a certain extent, low-level acoustic properties also account for self-voice recognition. Without any a priori hypotheses, we placed our

participants' voices in other-centered and gender-matched voice spaces(Baumann & Belin, 2010) and observed a correlation between acoustic distances and discriminability ratings – however, only for male participants. There might be several possible accounts such gender imbalance. First, female voices are less distinctive (Skuk & Schweinberger, 2013) and a lack of correlation might thus arise from a reduced variability in female voice distances, which was present in our data. Secondly, there is a right-hemisphere advantage for voice processing (Lattner, Meyer, & Friederici, 2005) and females are known to be less functionally lateralized than males (Boles, 2005). Finally, the difference might arise from the construction of voice spaces themselves. The second dimension of voice-spaces in Baumann and Belin's study differed for male and female voices – female voice-space contained lower formants (F1), whereas male voice-space higher formants (F5 – F4). Although both carry identity information (Latinus & Belin, 2012; Schweinberger et al., 2008), lower formants can be altered by the speaker (Maeda, 1990), whereas higher formants cannot, as they reflect the morphology of individual vocal cavities (Kitamura & Akagi, 1995). Higher formants were explicitly related to speaker identification (López et al., 2013) and were further shown to be advantageous specifically for self-voice recognition (Xu, Homae, Hashimoto, & Hagiwara, 2013). It is thus possible that SOVD depends on acoustic differences between self and other voices only for males, because male voice space contains higher formant information. Our data further supports the specificity of higher formants to self-voice processing by showing that increase in high-formant distances increases SOVD, but not familiar-other discrimination. This difference in reliance on higher formants for self- and familiar voices seems to be further increased with bone compared to air conduction. However, these observations are purely correlational and more work is needed to elucidate both the gender differences in SOVD and the relationship between self-voice perception and higher formants. In sum, these findings show that both familiarity and acoustic processing contribute to SOVD and future studies should identify ways to delineate the corresponding contributions of these factors.

Besides approximating experimental self-voices to their internal representation, another important novelty of this work is the sensitivity of the proposed task that enables pinpointing perceptual specificities in SOVD, while avoiding ceiling effects accompanying recent self-voice studies (Candini et al., 2018; Candini et al., 2014; Hughes & Nicholson, 2010; Rosa et al., 2008). Specifically, we observed that multimodal self-voice presentation is most advantageous for other-dominant self-other voice morphs, suggesting that rather than labeling an ambiguous voice as 'self', bone conduction facilitates discarding an ambiguous voice as being 'not self'. This opens the door to clinical investigations of etiology of auditory-verbal hallucinations, which are specifically thought to arise due to erroneous mislabeling of internal subvocalizations as being of 'not self' origin (Frith, 1992; Frith & Done, 1989). However, it should be pointed out that most participants spontaneously complained about the SOVD task being very difficult and believed they were not able to perform it, even if they actually performed well. Accordingly, in Study 1 we observed big differences in performance across participants – some even inverted self and other voices (i.e. yielded psychometric curves with negative slopes), and, interestingly, some could perform the task well only with one sound conduction type. That is why, in Study 3, we introduced a self-recognition task at the end of experiment, with the purpose of narrowing the SOVD analysis down only to those participants who were able to recognize their own voice without voice morphing. To our surprise, nine out of 52 (17.3%) participants could not recognize their unmorphed voice in more than half of self-recognition task's trials. This shows that recognizing own voice in short vocalizations (even without voice morphing) is not as trivial as it might seem, although it is shown to suffice for speaker identification (Zarate, Tian, Woods, & Poeppel, 2015).

To conclude, the impact of the work presented here is both methodological and scientific. For one, we propose a new perspective on addressing self-voice-related research questions by demonstrating a more ecological method for auditory self-identification. Based on these findings, future studies should avoid presenting self-voice stimuli through traditional air-conducting media, especially considering the increasing availability of bone-conduction headsets. For another, we shed new light on the phenomenology of the self by portraying self-voice as a fundamentally multimodal construct. Future imaging studies should identify neural underpinnings of multimodally presented self-voice.

2.1.6 References

Adler, D., Herbelin, B., Similowski, T., & Blanke, O. (2014). Breathing and sense of self: Visuo-respiratory conflicts alter body self-consciousness. Respiratory Physiology & Neurobiology, 203, 68–74. https://doi.org/10.1016/j.resp.2014.08.003

Allen, P. P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M., Johns, L. C., & McGuire, P. K. (2005). Neural correlates of the misattribution of self-generated speech. Human Brain Mapping, 26(1), 44–53. https://doi.org/10.1002/hbm.20120

Allen, P. P., Johns, L. C., Fu, C. H. Y., Broome, M. R., Vythelingum, G. N., & McGuire, P. K. (2004). Misattribution of external speech in patients with hallucinations and delusions. Schizophrenia Research, 69(2–3), 277–287. https://doi.org/10.1016/j.schres.2003.09.008

Alzueta, E., Melcón, M., Poch, C., & Capilla, A. (2019). Is your own face more than a highly familiar face? Biological Psychology, 142(January), 100–107. https://doi.org/10.1016/j.biopsycho.2019.01.018

Ameller, A., Dereux, A., Dubertret, C., Vaiva, G., Thomas, P., & Pins, D. (2015). 'What is more familiar than I? Self, other and familiarity in schizophrenia. Schizophrenia Research, 161(2–3), 501–505. https://doi.org/10.1016/j.schres.2014.12.007

Aspell, J. E., Heydrich, L., Marillier, G., Lavanchy, T., Herbelin, B., & Blanke, O. (2013). Turning Body and Self Inside Out: Visualized Heartbeats Alter Bodily Self-Consciousness and Tactile Perception. Psychological Science, 24(12), 2445–2453. https://doi.org/10.1177/0956797613498395

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software. https://doi.org/10.18637/jss.v067.i01

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. Psychological Research, 74(1), 110–120. https://doi.org/10.1007/s00426-008-0185-z

Békésy, G. V. (1949). The Structure of the Middle Ear and the Hearing of One's Own Voice by Bone Conduction. Journal of the Acoustical Society of America, 21(3), 217–232. https://doi.org/10.1121/1.1906501

Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. Neuroscience and Biobehavioral Reviews, 47, 717–734. https://doi.org/10.1016/j.neubiorev.2014.10.022

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. Trends in Cognitive Sciences, 13(1), 7–13. https://doi.org/10.1016/j.tics.2008.10.003

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness. Neuron, 88(1), 145–166. https://doi.org/10.1016/j.neuron.2015.09.029

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.12. Retrieved April 13, 2020, from http://www.praat.org/

Boles, D. B. (2005). A large-sample study of sex differences in functional cerebral lateralization. Journal of Clinical and Experimental Neuropsychology, 27(6), 759-768. https://doi.org/10.1081/13803390590954263

Candini, M., Avanzi, S., Cantagallo, A., Zangoli, M. G., Benassi, M., Querzani, P., … Frassinetti, F. (2018). The lost ability to distinguish between self and other voice following a brain lesion. NeuroImage: Clinical, 18(March), 903–911. https://doi.org/10.1016/j.nicl.2018.03.021

Candini, Michela, Zamagni, E., Nuzzo, A., Ruotolo, F., Iachini, T., & Frassinetti, F. (2014). Who is speaking? Implicit and explicit self and other voice recognition. Brain and Cognition, 92, 112–117. https://doi.org/10.1016/j.bandc.2014.10.001

Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. PLoS ONE, 10(4), 1–12. https://doi.org/10.1371/journal.pone.0121945

Dobrev, I., Sim, J. H., Stenfelt, S., Ihrle, S., Gerig, R., Pfiffner, F., … Röösli, C. (2017). Sound wave propagation on the human skull surface with bone conduction stimulation. Hearing Research, 355, 1–13. https://doi.org/10.1016/j.heares.2017.07.005

Douglas, W., & Gibbins, K. (1983). Inadequacy of voice recognition as a demonstration of self-deception. Journal of Personality and Social Psychology, 44(3), 589–592. https://doi.org/10.1037/0022-3514.44.3.589

Emami, S. F., Pourbakht, A., Sheykholeslami, K., Kamali, M., Behnoud, F., & Daneshi, A. (2012). Vestibular Hearing and Speech Processing. ISRN Otolaryngology, 2012, 1–7. https://doi.org/10.5402/2012/850629

Frith, C. D. (1992). The cognitive neuropsychology ofschizophrenia. Erlbaum. https://doi.org/10.1016/0028-3932(93)90120-O

Frith, Christopher D., & Done, D. J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. Psychological Medicine, 19(2), 359-363. https://doi.org/10.4324/9781315630502

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. Trends in Cognitive Sciences, 4(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., & Bruneau, N. (2015). Is my voice just a familiar voice? An electrophysiological study. Social Cognitive and Affective Neuroscience, 10(1), 101–105. https://doi.org/10.1093/scan/nsu031

Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. Journal of Personality and Social Psychology, 37(2), 147–169. https://doi.org/10.1037/0022-3514.37.2.147

Harkavy-Friedman, J. M., Kimhy, D., Nelson, E. A., Venarde, D. F., Malaspina, D., & Mann, J. J. (2003). Suicide attempts in schizophrenia: the role of command auditory hallucinations for suicide. The Journal of Clinical Psychiatry., 64(8), 871–874.

Holzman, P. S., & Rousey, C. (1966). The voice as a percept. Journal of Personality and Social Psychology, 4(1), 79–86. https://doi.org/10.1037/h0023518

Holzman, P. S., Rousey, C., & Snyder, C. (1966). On listening to one's own voice: Effects on psychophysiological responses and free associations. Journal of Personality and Social Psychology, 4(4), 432–441. https://doi.org/10.1037/h0023790

Hughes, S. M., & Nicholson, S. E. (2010). The processing of auditory and visual recognition of self-stimuli. Consciousness and Cognition, 19(4), 1124–1134. https://doi.org/10.1016/j.concog.2010.03.001

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. Proceedings of the National Academy of Sciences, 106(4), 1245–1248. https://doi.org/10.1073/pnas.0810063106

Jalenques, I., Enjolras, J., & Izaute, M. (2013). Valence émotionnelle des mots. L'Encéphale, 39, 189–197.

Jeannerod, M. (2003). The mechanism of self-recognition in humans. Behavioural Brain Research, 142, 1–15. https://doi.org/10.1016/S0166-4328(02)00384-4

Kawahara, H., Morise, M., Banno, H., & Skuk, V. G. (2013). Temporally variable multi-aspect N-way morphing based on interference-free speech representations. In 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2013 (pp. 1–10). IEEE. https://doi.org/10.1109/APSIPA.2013.6694355

Kimura, M., & Yotsumoto, Y. (2018). Auditory traits of "own voice." PLoS ONE, 13(6), 1–16. https://doi.org/10.1371/journal.pone.0199443

Kitamura, T., & Akagi, M. (1995). Speaker individualities in speech spectral envelopes. Journal of the Acoustical Society of Japan (E) (English Translation of Nippon Onkyo Gakkaishi), 16(5), 283–289. https://doi.org/10.1250/ast.16.283

Kleiner, M., Brainard, D. H., Pelli, D. G., Ingling, A., Murray, R., & Broussard, A Ingling, R Murray, C. (2007). What's new in Psychtoolbox-3? Perception, 36(14), 1–16. https://doi.org/10.1068/v070821

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Latinus, M., & Belin, P. (2012). Perceptual auditory aftereffects on voice identity using brief vowel stimuli. PLoS ONE, 7(7), 1–7. https://doi.org/10.1371/journal.pone.0041384

Latinus, M., McAleer, P., Bestelmeyer, P. E. G., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex: Supplemental information. Current Biology, 23(12), 1075–1080. https://doi.org/10.1016/j.cub.2013.04.055

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. Human Brain Mapping, 24(1), 11–20. https://doi.org/10.1002/hbm.20065

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video Ergo Sum: Manipulating Bodily Self-Consciousness. Science, 317(5841), 1096–1099. https://doi.org/10.1126/science.1143439

López, S., Riera, P., Assaneo, M. F., Eguía, M., Sigman, M., & Trevisan, M. A. (2013). Vocal caricatures reveal signatures of speaker identity. Scientific Reports, 3, 1–7. https://doi.org/10.1038/srep03407

Lüdecke, D. (2018). sjPlot: Data Visualization for Statistics in Social Science. R package version 2.6.2. https://doi.org/10.5281/zenodo.1308157

Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In Speech production and speech modelling (pp. 131–149). Springer, Dordrecht.

Manning, C., Mermagen, T., & Scharine, A. (2017). The effect of sensorineural hearing loss and tinnitus on speech recognition over air and bone conduction military communications headsets. Hearing Research, 349, 67–75. https://doi.org/10.1016/j.heares.2016.10.019

Maurer, D., & Landis, T. (1990). Role of bone conduction in the self-perception of speech. Folia Phoniatrica et Logopaedica, 42(5), 226–229. https://doi.org/10.1159/000266070

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., … Kojima, S. (2001). Neural substrates for recognition of familiar voices: A PET study. Neuropsychologia, 39(10), 1047–1054. https://doi.org/http://dx.doi.org/10.1016/S0028-3932%2801%2900037-9

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain- A meta-analysis of imaging studies on the self. NeuroImage, 31(1), 440–457. https://doi.org/10.1016/j.neuroimage.2005.12.002

Olivos, G. (1967). Response delay, psychophysiologic activation, and recognition of one's own voice. Psychosomatic Medicine, 29(5), 433–440. https://doi.org/10.1097/00006842-196709000-00003

Orepic, P., Bernasconi, F., Faggela, M., Faivre, N., & Blanke, O. (2020). Robotically-mediated sensorimotor stimulation induces identity-specific auditory-verbal hallucinations in healthy individuals. In Preparation.

Orepic, Pavo, Rognini, G., Kannape, O. A., Faivre, N., & Blanke, O. (2020). Sensorimotor conflicts induce somatic passivity and louden quiet voices in healthy listeners. BioRxiv. https://doi.org/https://doi.org/10.1101/2020.03.26.005843

Park, H. D., & Blanke, O. (2019). Coupling Inner and Outer Body for Self-Consciousness. Trends in Cognitive Sciences, 23(5), 377–388. https://doi.org/10.1016/j.tics.2019.02.002

Pearson, K., & Filon, L. N. G. (1898). Mathematical Contributions to Theory of Evolution: IV. On the Probable Errors of Frequency Constants and on the Influence of Random Selection and Correlation. Proceedings of the Royal Society of London, 62(191), 229–311. https://doi.org/10.1098/rsta.1898.0007

Pfeiffer, C., Schmutz, V., & Blanke, O. (2014). Visuospatial viewpoint manipulation during full-body illusion modulates subjective first-person perspective. Experimental Brain Research, 232(12), 4021–4033. https://doi.org/10.1007/s00221-014-4080-0

Pörschmann, C. (2000). Influences of bone conduction and air conduction on the sound of one's own voice. Acustica, 86(6), 1038–1045.

Preston, C., Kuper-Smith, B. J., & Ehrsson, H. H. (2016). Owning the body in the mirror: The effect of visual perspective and mirror view on the full-body illusion. Scientific Reports, 5(1), 18345. https://doi.org/10.1038/srep18345

Qin, P., Wang, M., & Northoff, G. (2020). Linking bodily, environmental and mental states in the self — A three-level model based on a meta-analysis. Neuroscience and Biobehavioral Reviews, 115, 77–95. https://doi.org/10.1016/j.neubiorev.2020.05.004

R Core Team. (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Reinfeldt, S., Östli, P., Håkansson, B., & Stenfelt, S. (2010). Hearing one's own voice during phoneme vocalization—Transmission by air and bone conduction. The Journal of the Acoustical Society of America, 128(2), 751–762. https://doi.org/10.1121/1.3458855

Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P., & Belin, P. (2008). Investigations of hemispheric specialization of self-voice recognition. Brain and Cognition, 68(2), 204–214. https://doi.org/10.1016/j.bandc.2008.04.007

Rousey, C., & Holzman, P. S. (1967). Recognition of One'S Own Voice. Journal of Personality and Social Psychology, 6(4), 464–466. https://doi.org/10.1037/h0024837

Schuerman, W. L., Meyer, A., & McQueen, J. M. (2015). Do we perceive others better than ourselves? A perceptual benefit for noise-vocoded speech produced by an average speaker. PLoS ONE, 10(7), 1–18. https://doi.org/10.1371/journal.pone.0129731

Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N., … Zäske, R. (2008). Auditory Adaptation in Voice Perception. Current Biology, 18(9), 684–688. https://doi.org/10.1016/j.cub.2008.04.015

Shuster, L. I. (1998). The perception of correctly and incorrectly produced /r/. Journal of Speech, Language, and Hearing Research, 41(4), 941–950.

Shuster, L. I., & Durrant, J. D. (2003). Toward a better understanding of the perception of self-produced speech. Journal of Communication Disorders, 36(1), 1–11. https://doi.org/10.1016/S0021-9924(02)00132-6

Skuk, V. G., & Schweinberger, S. R. (2013). Gender differences in familiar voice identification. Hearing Research, 296, 131–140. https://doi.org/10.1016/j.heares.2012.11.004

Stenfelt, S. (2011). Acoustic and physiologic aspects of bone conduction hearing. Implantable Bone Conduction Hearing Aids, (71), 10–21. https://doi.org/10.1159/000323574

Stenfelt, S. (2016). Model predictions for bone conduction perception in the human. Hearing Research, 340, 135–143. https://doi.org/10.1016/j.heares.2015.10.014

Stenfelt, S., & Goode, R. L. (2005). Bone-conducted sound: Physiological and clinical aspects. Otology and Neurotology, 26(6), 1245–1261. https://doi.org/10.1097/01.mao.0000187236.10842.d5

Sugiura, M., Kawashima, R., Nakamura, K., Okada, K., Kato, T., Nakamura, A., … Fukuda, H. (2000). Passive and active recognition of one's own face. NeuroImage, 11(1), 36–48. https://doi.org/10.1006/nimg.1999.0519

Sui, J., & Humphreys, G. W. (2017). The ubiquitous self: What the properties of self-bias tell us about the self. Annals of the New York Academy of Sciences, 1396(1), 222–235. https://doi.org/10.1111/nyas.13197

Todd, N. P. M. A., Cody, F. W. J., & Banks, J. R. (2000). A saccular origin of frequency tuning in myogenic vestibular evoked potentials?: implications for human responses to loud sounds. Hearing Research, 141(1–2), 180–188. https://doi.org/10.1016/S0378-5955(99)00222-1

Tonndorf, J. (1976). Bone conduction. In Auditory System (pp. 37–84). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-66082-5_13

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. Nature, 423(6942), 866–869. https://doi.org/10.1038/nature01710

Uddin, L. Q., Kaplan, J. T., Molnar-Szakacs, I., Zaidel, E., & Iacoboni, M. (2005). Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: An event-related fMRI study. NeuroImage, 25(3), 926–935. https://doi.org/10.1016/j.neuroimage.2004.12.018

Vurma, A. (2014). The timbre of the voice as perceived by the singer him-/herself. Logopedics Phoniatrics Vocology, 39(1), 1–10. https://doi.org/10.3109/14015439.2013.775334

Wheatstone, C. (1827). Experiments on audition. Quarterly Journal of Science, Literature and Art, 24, 67–72.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. Springer-Verlag New York. Retrieved from http://ggplot2.org

Won, S. Y., Berger, J., & Slaney, M. (2014). Simulation of One's Own Voice in a Two-parameter Model. Proceedings of the International Conference on Music Perception and Cognition (ICMPC).

Xu, M., Homae, F., Hashimoto, R., & Hagiwara, H. (2013). Acoustic cues for the recognition of self-voice and other-voice. Frontiers in Psychology, 4(October), 1–7. https://doi.org/10.3389/fpsyg.2013.00735

Zarate, J. M., Tian, X., Woods, K. J. P., & Poeppel, D. (2015). Multiple levels of linguistic and paralinguistic features contribute to voice recognition. Scientific Reports, 5, 11475. https://doi.org/10.1038/srep11475

2.1.7 Supplementary material

2.1.7.1 Words in Study 4

Nine one-syllable French words (translated to English: nail, whip, ax, blade, fight, bone, rat, blood, saw, worm) were chosen from the list of 100 negatively-valenced words, as rated by 20 schizophrenic patients and 97 healthy participants (Jalenques, Enjolras, & Izaute, 2013). Negative words were chosen to match experimental paradigms of our other studies (Orepic, Bernasconi, Faggela, Faivre, & Blanke, 2020; Orepic, Rognini, Kannape, Faivre, & Blanke, 2020).

2.1.7.2 Reaction times

Both in studies 1 and 2, the linear mixed-effect regression on response times identified the main effect of Previous Exposure (1: estimate=-2.03, t(3574)=-14.4, p<.001; 2: estimate=-0.38, t(15.98)=-2.99, p=.009). Participants responded faster when they previously heard the target voice. In none of the three tasks of Study 3 (self-unfamiliar, familiar-unfamiliar, self-familiar) were reaction times significantly affected by the effect of Conduction, nor was there a significant interaction with the effect of Voice Morph. Polynomial expansion of Voice Morph was significant for all tasks, indicating the 'inverse-u' shape of reaction times with the increasing levels of Voice Morph. The details of the model are given in Table 1 and the results are illustrated on Figure 12.

Table 1. Model details for Reaction Times.
An overview of the linear mixed-effects models with Reaction Times as dependent variable for all tasks of Study 3.

Task	Effect	estimate	df	t value	p value
Self-unfamiliar	Conduction	0.03	42.86	0.69	0.492
	Voice Morph	-1.92	4718.88	-3.38	< 0.001
	Conduction * Voice Morph	-0.15	4718.68	-0.18	0.854
Familiar-unfamiliar	Conduction	0.04	42.98	1.34	0.188
	Voice Morph	-3.73	4825.44	-6.49	< 0.001
	Conduction * Voice Morph	0.66	4825.6	0.81	0.417
Self-familiar	Conduction	0.05	43	1.19	0.239
	Voice Morph	-2.17	4698.25	-3.75	< 0.001
	Conduction * Voice Morph	0.87	4697.89	1.07	0.287
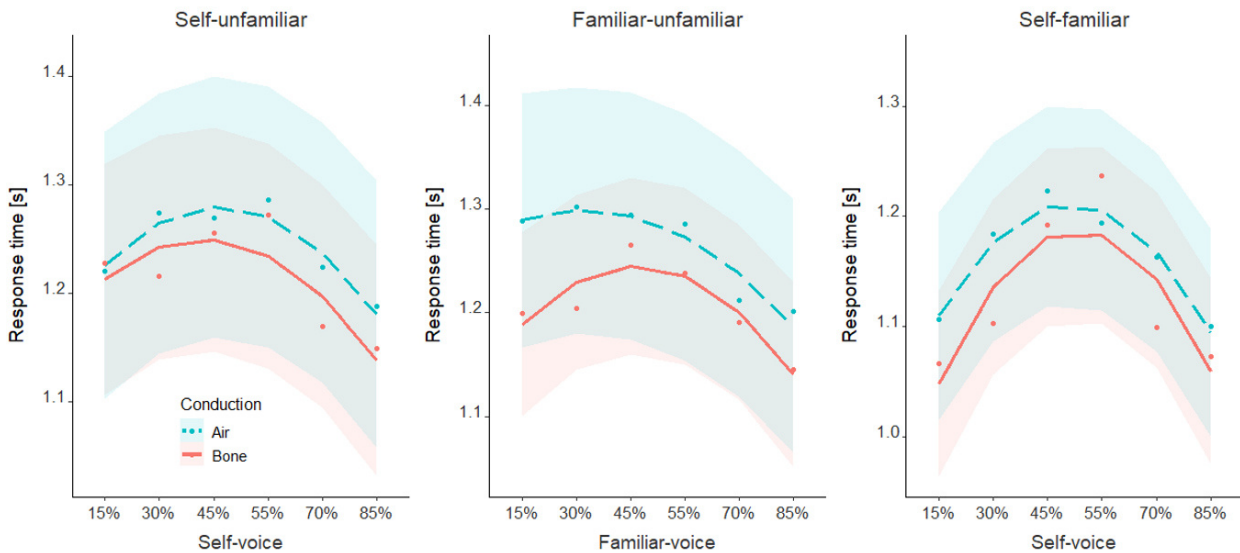
Figure 12. Reaction times for the three tasks of Study 3.
Shaded areas represent 95% confidence intervals. Sound conduction did not affect reaction times in any task.

2.1.7.3   Self-confusion

Pearson and Filon's z-test did not indicate significant differences in correlations between miss and familiar-FA rates with bone and air conduction (z=-0.74, p=.462, Figure 13, left). No such differences were observed for unfamiliar-FAs (z=1, p=.314, Figure 13, right).
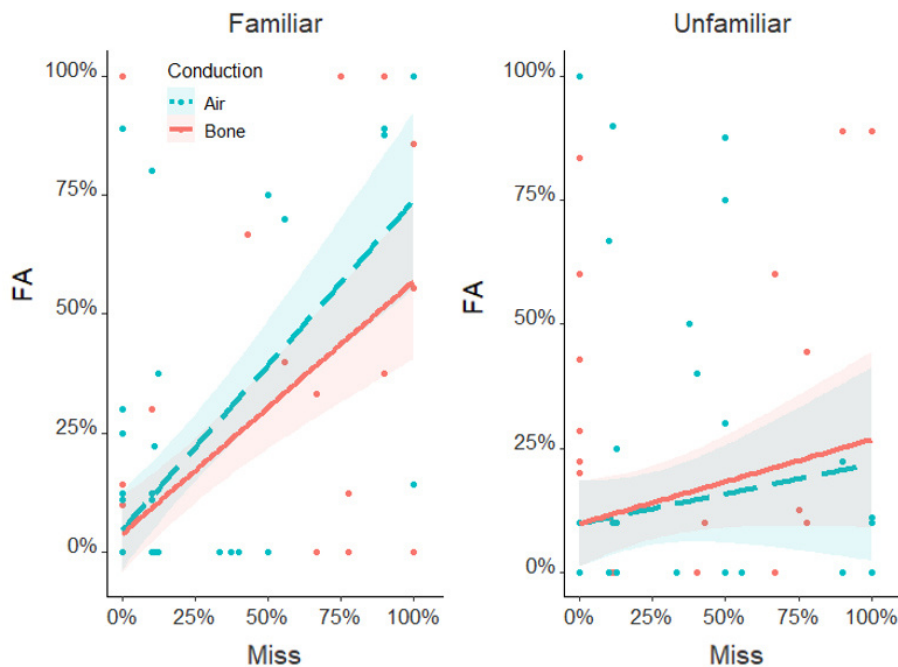
Figure 13. Self-confusion effects were unaffected by sound conduction.
Regression plots indicate relationships between false alarm rates for familiar (left) and unfamiliar (right) voice with the miss rate for self-voice, separately for air and bone conduction. Shaded areas around linear regressions indicate 95% confidence intervals. Neither familiar-FA nor unfamiliar-FA rates were affected by sound conduction.

2.1.7.4 Prototype-based coding of self-voices

We conducted two exploratory analyses investigating the effects of sound conduction and familiarity of other voice on the relationship between voice-space distances and task performance. None were significant (all $p > 0.05$ in Pearson and Filon's z-test) (Figure 14).
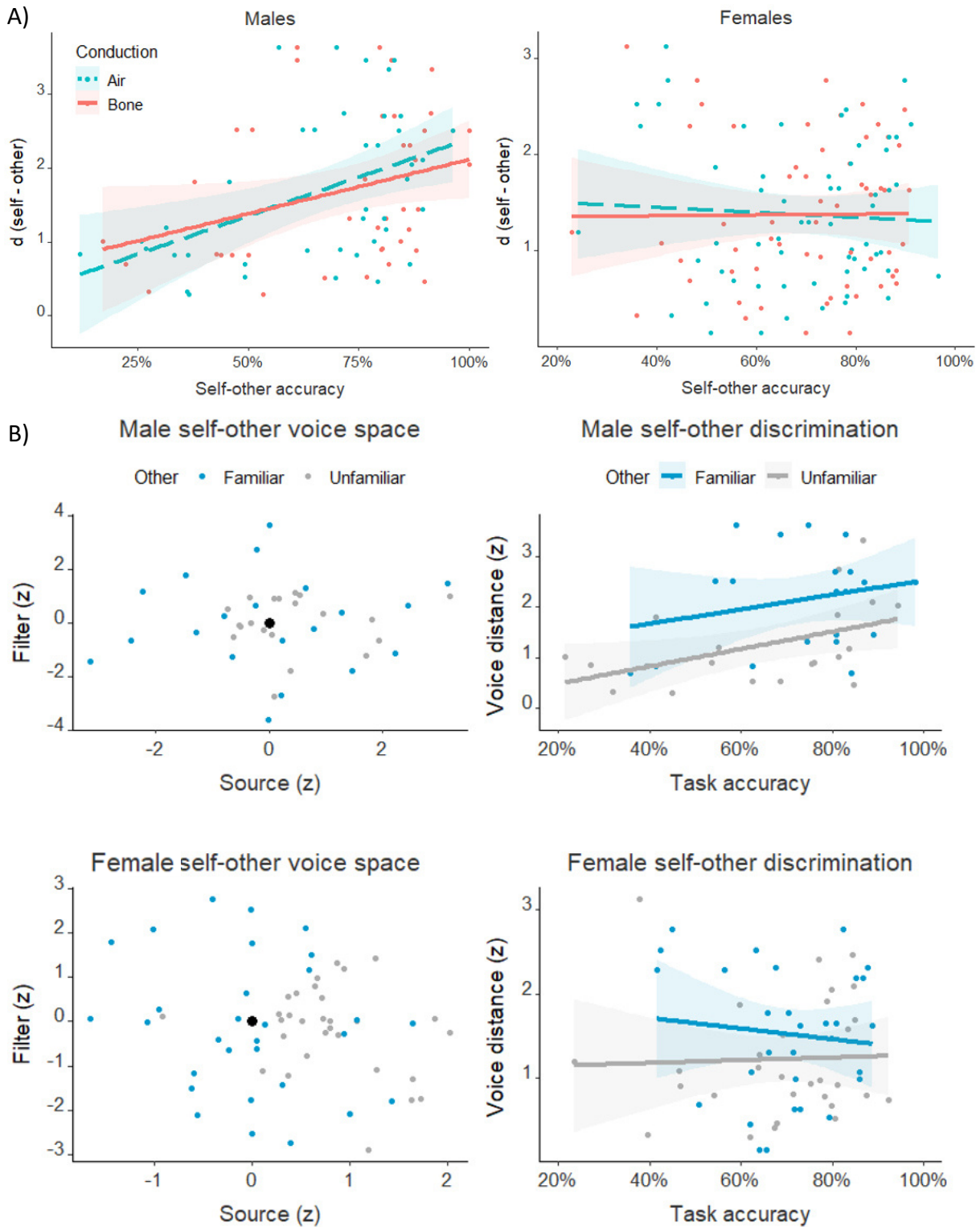
Figure 14. Effects of sound conduction and familiarity on the voice-space analysis.

Effects of sound conduction and familiarity on voice-space analysis. Neither sound conduction (A) nor other-voice familiarity (B) did affect the relationship between task performance and voice-space distances.

2.2 Study 2: EEG spatiotemporal patterns underlying self-other voice discrimination

Pavo Orepic[1,*], Giannina Rita Iannotti[2,3,*], Sixto Alcoba-Banqueri[1], Dorian Fernand Alfred Garin[3], Karl Schaller[3], Olaf Blanke[1,4], Christoph Michel[2]

1 Laboratory of Cognitive Neuroscience, Center for Neuroprosthetics and Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Switzerland

2 Functional Brain Mapping Lab, Department of Fundamental Neurosciences, University of Geneva, Switzerland

3 Department of Neurosurgery, Geneva University Medical Center & Faculty of Medicine, University of Geneva, Switzerland

4 Faculty of Medicine, University of Geneva, Geneva, Switzerland

* The two authors contributed equally to the conception of study, analyses and redaction of the manuscript

Corresponding author

Olaf Blanke

Bertarelli Chair in Cognitive Neuroprosthetics, Center for Neuroprosthetics & Brain Mind Institute, School of Life Sciences, Campus Biotech, Swiss Federal Institute of Technology (EPFL), 1012 Geneva, Switzerland

E-mail: olaf.blanke@epfl.ch

Tel: +41 (0)21 693 69 21

Fax: +41 (0)21 693 69 22

2.2.1 Abstract

There is growing evidence showing that the representation of the self recruits special or privileged systems across different functions and modalities. Compared to self-face and self-body representations, surprisingly few studies have investigated neural underpinnings specific to self-voice. Moreover, self-voice stimuli in those studies were consistently presented through air and lacking bone conduction, thus rendering the sound of self-voice different to the self-voice heard during natural speech. Here, we combined psychophysics, voice-morphing technology, and high-density EEG in order to identify spatiotemporal patterns underlying self-other voice discrimination (SOVD), both with air- and bone-conducted stimuli. We identified a self-voice specific EEG topographic map occurring around 345 milliseconds after stimulus onset that activated an extended network involving insula, cingulate cortex, and medial temporal lobe structures. Occurrence of this map was modulated both with SOVD task performance and bone conduction. Specifically, the better participants were at discriminating their own from a stranger's voice, the less frequently they activated this network. In addition, the same network was recruited less frequently with bone conduction, which, accordingly, increased the SOVD task performance. We are the first to directly associate behavioral and neural mechanisms underlying self-voice perception, thereby identifying the modulatory effects of bone conduction. Furthermore, this work is of vast clinical relevance, as it reveals neural correlates of SOVD impairments, that are believed to account for auditory-verbal hallucinations, the most common and highly distressing psychiatric symptom.

Keywords

Self-voice, self-other voice discrimination, EEG, EEG segmentation, bone conduction, insula, cingulate cortex

2.2.2 Introduction

Identifying individuals is an important cognitive ability laying a foundation for building social interactions and establishing personal relationships. Although face is the main carrier of human identity (Blank, Wieland, & von Kriegstein, 2014), people can unequivocally be identified also through their voice (Belin, Fecteau, & Bédard, 2004). Furthermore, inferring information about person's identity seems to follow similar processing steps in both modalities (Yovel & Belin, 2013). Whether the process of identifying oneself is equivalent to identifying other (highly) familiar individuals is still a question of debate (Gillihan & Farah, 2005). The self could be defined as an entity that is distinct from the environment and other humans to which certain mental events and actions are ascribed (Kircher & David, 2003). The special nature of the self is thought to arise from multisensory integration of bodily signals (Blanke, 2012; Blanke, Slater, & Serino, 2015; Park & Blanke, 2019; Tsakiris, 2017) and sensorimotor congruency resulting from the interactions of the body with the environment (Blakemore, Wolpert, & Frith, 2000; Braun et al., 2018; Kannape & Blanke, 2012; Tsakiris & Haggard, 2005). Accordingly, self-specificity has been reported in many neuroimaging studies that investigated self-referential processes across different functional domains (e.g. emotional, spatial, memory) (Northoff et al., 2006). Compared to the extensive work done on self-face representation (for an overview see Uddin, Kaplan, Molnar-Szakacs, Zaidel, & Iacoboni, 2005), providing considerable evidence about the special nature of the self in the visual modality (Alzueta, Melcón, Poch, & Capilla, 2019), self-voice representation has been investigated to a surprisingly lesser extent, and no such claim could yet be made. However, acquiring a better understanding of neural mechanisms underlying self-voice perception is of utmost importance, as erroneous self-voice misattribution is thought to account for auditory-verbal hallucinations (Ford & Mathalon, 2005; Frith, 1992; Frith & Done, 1989; Shergill et al., 2014), the most common hallucination in schizophrenia, associated with high degrees of distress in the affected population (Harkavy-Friedman et al., 2003).

To date, only a few neuroimaging studies investigated self-voice-specific activations and yielded with inconsistent findings. Electroencephalographic (EEG) investigations of self-voice specificity have mostly been confined to the analysis of single-electrode evoked responses following self and other voices. Specifically, P3 component in an oddball paradigm was observed to be lower for self-voice, compared to unfamiliar (Graux et al., 2013) and familiar voices (Graux, Gomot, Roux, Bonnet-Brilhault, & Bruneau, 2015). (Conde, Gonçalves, & Pinheiro, 2016) additionally identified a later mismatch negativity (MMN) for self-voice and narrowed the P3 reduction down to experimental stimuli consisting of simple vocalizations. In a follow-up study (Conde, Gonçalves, & Pinheiro, 2018), however, the same authors reported a contradictory effect – i.e. an increased P3 amplitude for self-voices – and other authors failed to replicate the differences in P3 component between self and other voices (Liu, Li, Li, Lou, & Chen, 2019). In a PET study of (Nakamura et al., 2001), a contrast between self-voice and familiar voice activation peaked in the right inferior frontal sulcus and parainsular cortex. Similarly, right inferior frontal gyrus produced greater signal to self- compared to familiar voice in a

study of (Kaplan, Aziz-Zadeh, Uddin, & Iacoboni, 2008). (Allen et al., 2005) found that, when contrasted to unfamiliar voice, self-voice activity was associated to left inferior frontal and right anterior cingulate cortex. Together, these findings, although inconsistent, suggest that self-voice is represented differently from other voices. Importantly, however, no study correlated behavioral performance (e.g. ability to recognize own voice) to underlying neural activations.

Behavioral investigations that compared recognition of self-vs-other voices mostly showed lower accuracy or slower response times for self-voice compared to other-voice stimuli (Allen et al., 2005; Gur & Sackeim, 1979; Hughes & Nicholson, 2010; Rosa, Lassonde, Pinard, Keenan, & Belin, 2008; Schuerman, Meyer, & McQueen, 2015; Shuster, 1998), indicating an inability to recognize self-voice as well as other voices. This arguably reflects the lack of bone conduction in self-voice stimuli presented in corresponding studies, which is inevitably present while speaking (Békésy, 1949; Reinfeldt, Östli, Håkansson, & Stenfelt, 2010). Namely, when we speak, we hear our voice also through bone conduction, which applies a physical transformation to the sound of our voice (Stenfelt, 2016) and, besides auditory, often involves somatosensory (Ito, Tiede, & Ostry, 2009; Tremblay, Shiller, & Ostry, 2003) and vestibular (Emami et al., 2012; Todd, Cody, & Banks, 2000) processing. Therefore, a lack of bone conduction results in a discrepancy between the stimuli consistently presented in self-voice studies (i.e. air-conducted own voice recordings) and the internal self-voice representation, which is often accompanied with feelings of unpleasantness (Sackeim & Gur, 1978). Thus, the findings of the existing studies that identified neural underpinnings specific to hearing self-voice based on stimuli presented only through air conduction should be taken with caution. In our recent work (Orepic, Kannape, Faivre, & Blanke, in preparation), we showed that multimodal presentation of self-voice stimuli that augments audition with vibrotactile excitation facilitates self-voice recognition. Specifically, combining psychophysics with voice-morphing technology (Kawahara, Morise, Banno, & Skuk, 2013), we designed a sensitive self-other voice discrimination (SOVD) task that enables pinpointing of perceptual specificities underlying SOVD, and observed a better SOVD task performance with the stimuli presented through a commercial bone-conduction headset compared to traditional air conduction media (e.g. loudspeakers).

Here, we build up on those findings by investigating neural mechanisms underlying SOVD with a high-density EEG setup, both through air and bone conduction. We recorded evoked responses of healthy individuals hearing ambiguous self-other voice morphs (e.g. a voice morph could be created such that it contains 40% of self-voice and 60% of stranger's voice) and correlated neural activation with the ability to determine a dominant voice in such voice morphs. Crucially, compared to previous EEG investigations of self-voice that exploited traditional single-electrode analysis (Conde et al., 2016, 2018; Graux et al., 2015, 2013), here we applied a more thorough approach (Brunet, Murray, & Michel, 2011; Murray, Brunet, & Michel, 2008) that allows for identifying precise spatiotemporal characteristics of electric fields at the scalp as well as temporal dynamics of those fields. Importantly, this approach is completely reference-independent and thus renders statistically unambiguous results (Lehmann & Skrandies, 1980). We investigated dependency of various spatiotemporal EEG parameters (e.g. map duration, maximal global field potential, global explained variance, etc.) on hearing a varying degree of self-voice, its relationship to psychophysically-quantified SOVD performance, as well as potential additional effects of multimodal self-voice presentation through bone conduction. Finally, we identified underlying neural sources accounting for the process of discriminating own from a stranger's voice. Based on previous work, we expected to observe a self-voice specific EEG pattern in the late ERP components (Conde et al., 2016; Graux et al., 2015) that would project to prefrontal (Kaplan et al., 2008; Nakamura et al., 2001) and cortical midline structures (Allen et al., 2005; Northoff et al., 2006). Our main goal was, however, to identify a specific relationship of this pattern to behavioral SOVD performance and its dependence on bone conduction.

2.2.3 Materials and Methods

2.2.3.1 Participants

This study involved 26 healthy participants, 14 female, mean age ± SD: 37.4 ± 14.7 years old. All participants were right-handed, reported no hearing deficits and no history of psychiatric or neurological disorders. They were instructed on the conduct of the task, gave informed consent in accordance with institutional guidelines (the Declaration of Helsinki and reference to the protocol PB_2016-01635, amendment 3 approved by the Commission Cantonal d' Ethique de la Recherche de Geneva) and received monetary compensation (CHF 20/h).

2.2.3.2 Self-other voice discrimination task

Prior to participating in the studies, participants' voices were recorded while vocalizing phoneme /a/ for approximately 1 to 2 seconds (Zoom H6 Handy recorder). Each recording was standardized for average intensity (-12 dBFS) and duration (500 milliseconds)

and cleaned from background noise (Audacity software). Short vocalizations were chosen to control for other linguistic and paralinguistic accounts for speaker identification, such as accent or prosody, constricting the identification process to simple acoustical properties of the voice. Previous work has shown that vocalizations suffice for speaker identification (Zarate, Tian, Woods, & Poeppel, 2015). Each participant's preprocessed voice was mixed with a target voice of a gender-matched unfamiliar person in order to generate voice morphs spanning a voice identity continuum between the two voices by using TANDEM-STRAIGHT (Kawahara, Morise, Banno, & Skuk, 2013), a voice morphing software package running in MATLAB. Six voice ratios (% self-voice: 15, 30, 45, 55, 70, 85) were chosen based on our previous work (Orepic et al., in preparation; Orepic, Rognini, Kannape, Faivre, & Blanke, 2020) and were repeated 10 times within a block in a randomized order (total of 60 trials per block). Inter-trial intervals jittered between 1 and 1.5 seconds to avoid predictability of stimulus onset.

Voice morphs were presented to participants either through bone-conducting headphones (Aftershokz Sports Titanium) or through laptop loudspeakers (air conduction). The order of air- and bone-conduction blocks was counterbalanced across participants. The study contained 10 experimental blocks, 5 of which were conducted consecutively with the same sound conduction type (air, bone), resulting in 50 trials per each voice-morph and sound conduction type. For each voice morph, participants were asked to indicate whether the voice they heard resembled their own or someone else's voice by clicking on a mouse button (2AFC). Importantly, participants were not presented with their unmorphed voice recordings prior to task execution, assuring that they performed the task by comparing voce morphs with the internal self-voice representation (Orepic et al., in preparation). The experimental paradigm was created in MATLAB 2017b with Psychtoolbox library (Kleiner et al., 2007).

During the task, electrophysiological (EEG) data were continuously recorded with a sampling rate of 1000 Hz using a 256-electrodes Hydrocel cap (Philips, Electrical Geodesics Inc.), referenced to the vertex (Cz). The impedance of electrodes was monitored carefully and kept below 40 kΩ. For the reference electrode, the impedance was kept below 10 kΩ. Bone-conducting headphones were installed on participants' heads underneath the EEG cap, by avoiding the overlap with any (namely temporal) electrodes. An illustration of the experimental paradigm and setup is given in Figure 15.
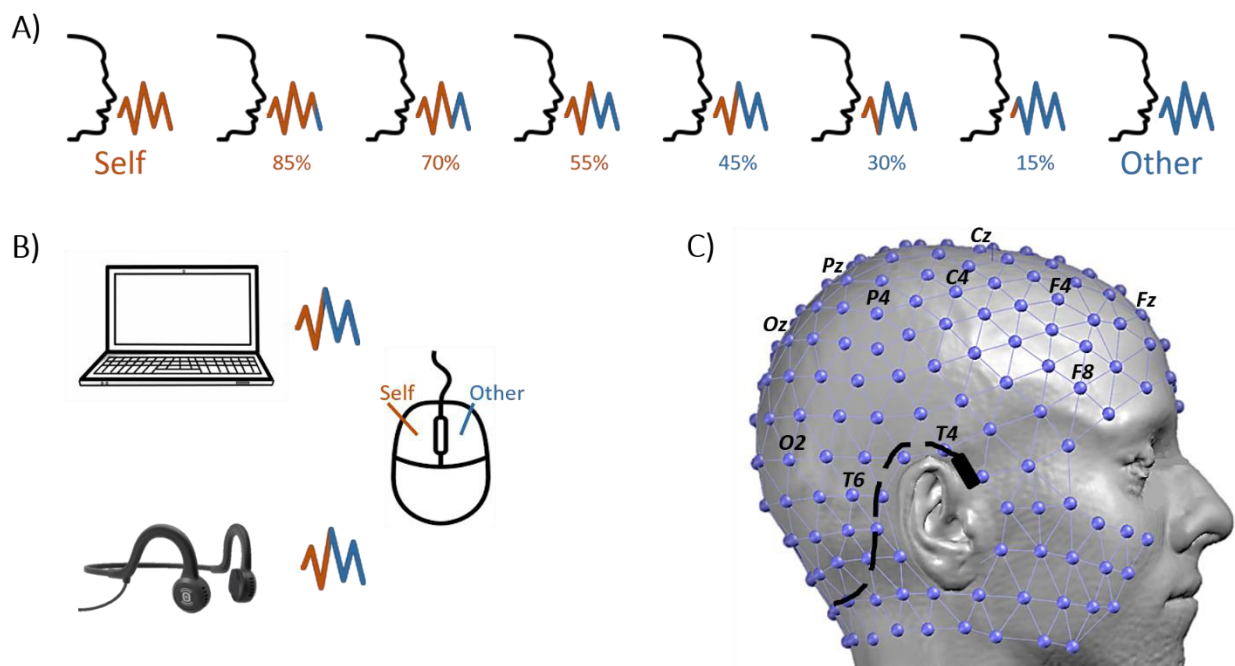
Figure 15. SOVD task.
A) Stimuli. Six self-other voice morphs between participant's voice (self, orange) and the voice of a gender-matched unfamiliar person (other, blue) were randomly presented 50 times throughout the experiment. **B)** Task. Voice morphs were presented either through air (laptop, above) or bone conduction (commercial headset, below). In every trial, participants responded whether the morph they hear resembles more to their or to someone else's voice by clicking on the corresponding mouse button. **C)** EEG setup. Bone conduction headphones (black) placed under a high-density EEG cap (lilac spheres and connections) formed by 256 electrodes organized as an extension of the standard clinical 10-20 setup (electrode names indicated in black).

2.2.3.3   EEG data pre-processing and selection of epochs

All the analyses of EEG data was performed with CARTOOL Software (Brunet et al., 2011). The EEG was reduced to 204 channels, by eliminating the electrodes covering the cheeks and the lowest neck surfaces because they were mostly not attached to participants' skin (due to individual anatomical configurations) and often contained motion artefacts (e.g. spontaneous chewing). First, the data were downsampled to 500 Hz and band-pass filtered between 1 and 40 Hz using non-causal Butterworth filters and a Notch filter of 50 Hz to eliminate (environmental) 50 Hz noise. Then, independent Component Analysis (ICA) was applied to remove eye-movement (eye blinks and saccades) and ECG artefacts, by using a MATLAB script based on the 'EEGlab runica' function (Bell & Sejnowski, 1995) that allows for component inspection both across time and scalp topographies. After ICA artefact removal, noisy electrodes were interpolated (3-D spherical spline, (Perrin, Pernier, Bertrand, & Echallier, 1989)) and a spatial filter was applied. Finally, for each participant, pre-stimulus epochs between -50 to 500 milliseconds were visually selected (i.e. they were excluded if residual artefacts, mostly due to motion, were observed) for each type of voice morph.

2.2.3.4   Data analysis

Statistical tests were performed with R (R Core Team, 2020), using the lme4 (Bates, Mächler, Bolker, & Walker, 2015), and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2018) packages. The results were illustrated using sjplot (Lüdecke, 2018) and ggplot2 (Wickham, 2016) packages.

2.2.3.4.1   Behavioral performance

Performance in self-other voice discrimination task was analyzed with mixed-effects binomial regressions with Accuracy as dependent variable and two fixed effects – Conduction (air, bone) and Morph (15, 30, 45, 55, 70, 85) – related with an interaction term. The Accuracy-variable indicated whether participants correctly identified the dominant voice in the presented morph. The model further contained the polynomial expansion of the Morph variable to level 2. Random effects included a by-participant random intercept, whereas by-participant random slopes for the fixed effects were added following model selection based on maximum likelihood. Trials with reaction times greater or smaller than two interquartile ranges from the median for each subject were considered as outliers and excluded. Equivalent mixed-effects regression was run for Response Time as a dependent variable.

2.2.3.4.2   ERPs segmentation

For each participant and each conduction type (air, bone) we grouped the ERPs belonging to each end of the self-other voice continuum, to increase the number of epochs from 50 to 100 and therefore the signal-to-noise ratio of the ERPs. Specifically, we averaged self-dominant (containing 85% and 70% self-voice) and other-dominant voice morphs (15% and 30% self-voice). In order to investigate the EEG topographic maps associated to the self- and other-dominant voices in the SOVD task, we grouped the corresponding ERPs across subjects and applied a group-averaged k-means cluster analysis (300 randomizations, 1-30 clusters, and a minimum cluster duration of 30 milliseconds). This procedure, also known as EEG segmentation (Brunet et al., 2011; Murray et al., 2008), was designed to identify the optimal number of EEG cluster maps that last for a predefined minimal period and that best describe the ERPs. The segmentation allows to define specific time windows and associated cluster maps to assess the (statistical) difference across subjects (Murray et al., 2008).

2.2.3.4.3   Back fitting

To assess the statistical difference between the self- and other-dominant voices, we projected back or "fitted back" the obtained maps on the ERPs of each subject, for each dominant morph and each conduction type. The back fitting was applied in three non-overlapping time windows selected from the group-averaged segmentation, by considering the associated cluster maps: i) [20-140] ms for maps 1 and 2; ii) [140-270] ms for map 3; iii) [270-500] ms for maps 4, 5, 6, and 7.

The back-fitting approach evaluates the spatial correlation between the cluster maps and the single ERP in each time point, by labelling the latter with the cluster map that shows the highest correlation. In our case we used a competitive attribution of the maps, (i.e., winner takes all, (Cabral, Kringelbach, & Deco, 2017; Fox & Raichle, 2007)), thereby taking into account the polarity of the maps. No label was assigned to the time points if the correlation between the ERP signal and any of all cluster maps was <0.25. The output of the back-fitting consisted in 12 parameters characterizing each cluster map for each subject, each dominant morph (self, other) and each conduction type (bone, air), (Table 2).

Table 2. Overview of back-fitting parameters.

Each parameter describes a property of an EEG cluster map for the pre-selected period.

Parameter	Description	Unit
Occurrence	The total number of data points in which a given map yields the highest spatial correlation value.	ms
Mean Duration	Mean number of *consecutive* data points in which a given map yields the highest spatial correlation value.	ms
First Occurrence	The first time point a given map yields the highest spatial correlation value.	ms
Last Occurrence	The last time point a given map yields the highest spatial correlation value.	ms
Max GFP	Maximal Global Field Power, or standard deviation of all electrodes, a measure of potential over time.	µV
Time of Max GFP	Occurrence of Maximal Global Field Power.	ms
Mean GFP	Mean Global Field Power.	µV
GEV	Global Explained Variance, i.e. how well a given map "explains" the data.	[0, 1]
Mean Correlation	Mean spatial correlation.	[0, 1]
Best Correlation	Highest spatial correlation.	[0, 1]
Time of Best Correlation	Occurrence of highest spatial correlation.	ms
GFP of Time of Best Correlation	GFP at the occurrence of highest spatial correlation.	µV

2.2.3.4.4   Analysis of map parameters

Considering that fitted parameters have a different range of values (e.g. GEV has values between 0 and 1, whereas Occurrence could have a value of a few hundred milliseconds), in order to be able to compare the experiment-driven differences in the values of different parameters, we first z-transformed all parameter values. Then, for each map, a linear mixed-effects regression with the dependent variable Parameter Value was performed with three fixed effects – Parameter (1-12), Conduction (air, bone) and Dominant Voice (self, other) – all related with an interaction term, and with by-participant random effects. Mixed-effects regression was used to account for possible missing values in the data, as for some participants, back-fitting procedure could omit certain maps. For the maps that indicated significant interactions involving the effect of Parameter, we ran additional mixed-effects regressions separately for each Parameter, directly assessing the effects of Conduction and Dominant Voice on Parameter Value of the corresponding parameter. This allowed us to investigate the effects of Conduction and Dominant Voice on all parameters while avoiding the problem of multiple comparisons.

2.2.3.4.5   EEG-behavior relationship

We further investigated the relationship between SOVD task performance (i.e. accuracy and response times) and parameters that were significantly affected by the effects of Dominant Voice and Condition. Thus, in the linear mixed-effects regression with the dependent variable Parameter Value and fixed effects Dominant Voice and Conduction, we added an additional effect of Accuracy, indicating the ratio of correct responses for the corresponding Dominant Voice and Conduction. All three effects were related with

an interaction term. The equivalent regression was run with the fixed effect of Response Time, indicating the average response time for the corresponding Dominant Morph and Conduction.

2.2.3.4.6 Source localization

To estimate the brain networks underlying the topographies of the scalp EEG resulting from the clustering procedure, we inverted the subjects' data for each map, dominant voice, and conduction type into the inverse space. We adapted the procedure that was previously applied for the EEG microstate and resting state analysis (Bréchet et al., 2019) to the analysis of ERPs data. For the forward model, we used a simplified realistic head model based on (Montreal Neurological Institute) MNI template, including the cerebellum, with consideration of skull thickness (Locally Spherical Model with Anatomical Constraints (LSMAC); for a review see (Michel & Brunet, 2019)) and a grid of around 6000 sources (solution points), distributed equally in the gray matter. The inverse solution was performed with LORETA (low-resolution brain electromagnetic tomography).

2.2.4 Results

Two participants were excluded from the analysis due to the poor quality of the EEG and six because they had a task performance lower than 50% independent of conduction type (air, bone). Similar number of participants was unable to perform the same task in our previous study (Orepic et al., in preparation). Our sample resulted therefore in 17 subjects (9 female, mean age ± SD: 37.3 ± 15.1 years old).

2.2.4.1 Behavioral results

A mixed-effects binomial regression in the self-other discrimination with Accuracy as dependent variable revealed a main effect of the polynomial expansion of the Morph variable (estimate=72.85, Z=17.18, p<0.001), indicating an u-shaped task performance with the increase of self-voice present in the morph. It further revealed a main effect of Conduction (estimate=-0.18, Z=2.93, p=0.003), reflected as higher accuracy for bone conduction. A two-way interaction between the effects of Conduction and Morph was also significant (estimate=19.13, Z=3.05, p=0.002), observed as a steeper slope for bone conduction. Post-hoc investigation of the interaction revealed that higher accuracy for bone conduction only in the morphs with least self-voice present (15%) (estimate=1.22, Z=2.57, p=0.01). Difference in accuracy between other voice morphs was not significant (all p>0.05). Together, participants discriminated own from a stranger's voice better with bone compared to air conduction and this was most prominent for other-dominant morphs (Figure 16, left).
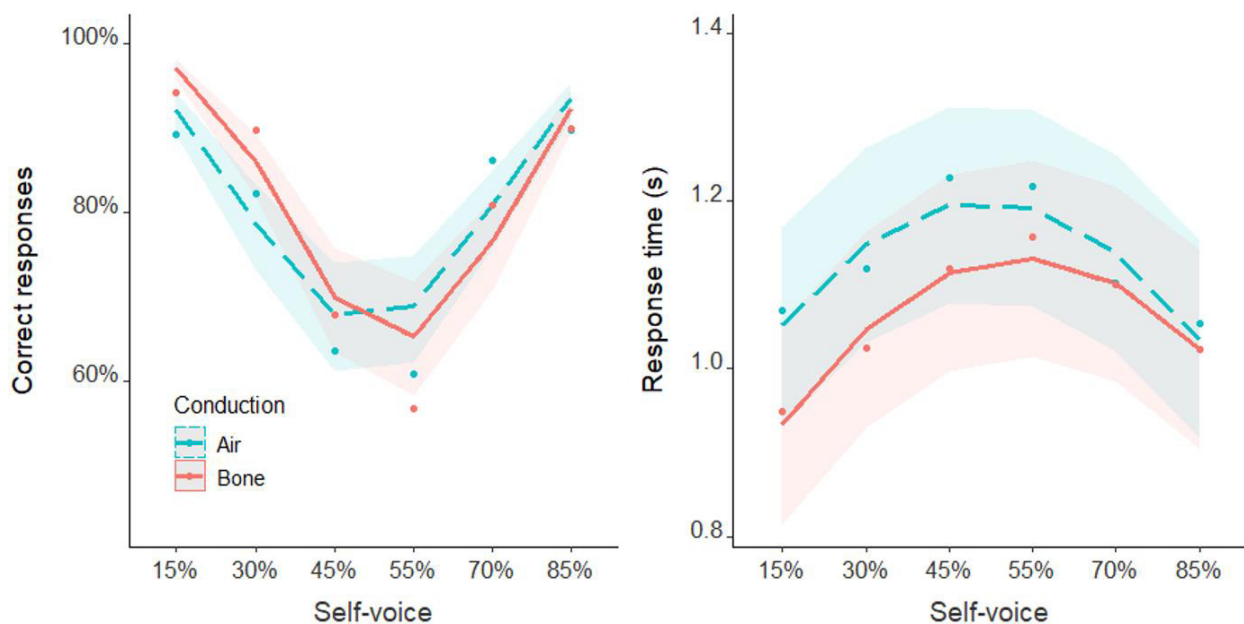
Figure 16. Behavioral results.
Behavioral results indicating the effects of the two forms of sound conduction (air, bone) on accuracy (left) and response times (right) in self-other voice discrimination task. The abscissa of both plots indicates the percentage of the self-voice present in a voice morph. The shaded areas around each curve represent the 95% confidence intervals. Accuracy was higher and responses were slower for bone conduction.

Linear mixed-effects regression for response times as dependent variable similarly revealed the main effect of the polynomial expansion of Morph variable (estimate=-6.05, t(9535)=-11.93, p<0.001), indicating a reversed u-shape of response times dependent on self-voice ratio. There was the main effect of Conduction (estimate=-0.07, t(9535)=-9.31, p<0.001), revealing faster response times for bone conduction. Interaction between the effects of Morph and Conduction was not significant (estimate=-0.2, t(9535)=0.28, p=0.783). Response times are shown at the right of Figure 16.

2.2.4.2 EEG results

2.2.4.2.1 Group average clustering

The k-means clustering executed on the group averaged ERPs for self-dominant and other-dominant voices and for both sound conduction types (air, bone) resulted in 7 different time segments and associated 7 cluster scalp EEG topographies, or maps (Figure 17). The investigation of missing values after the back-fitting procedure revealed a variable but relative low percentage of missing maps in our cohort for all the maps.
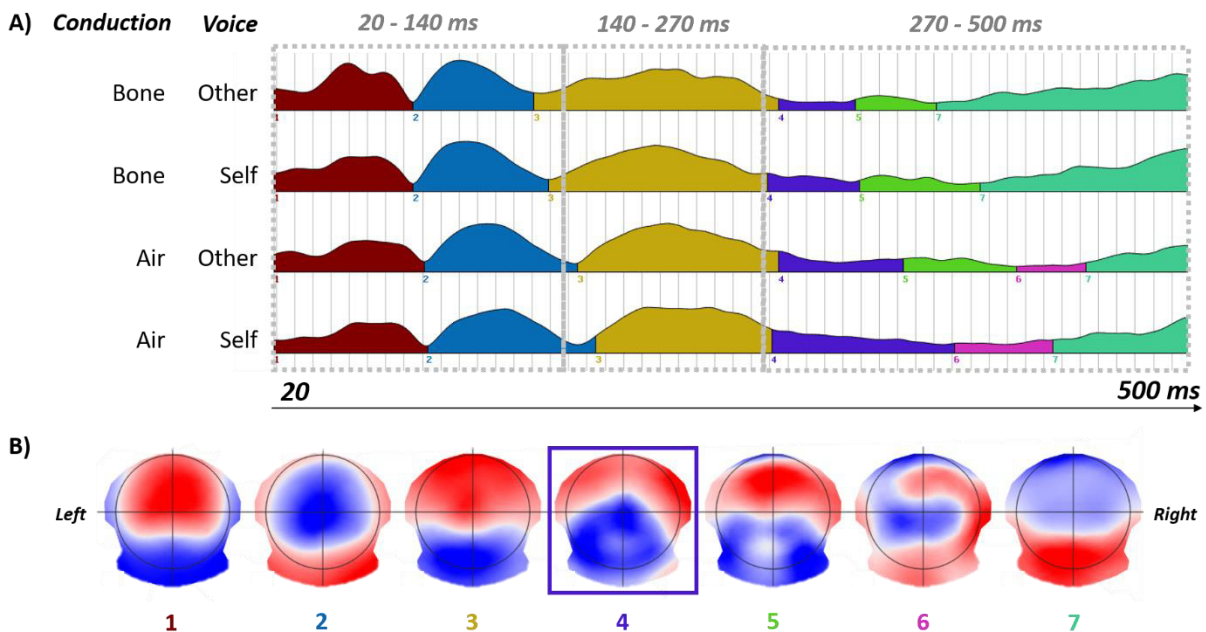
Figure 17. Results of the group-average ERP segmentation.

A) Different colors indicate different segments marked under the Global Field Power curves extracted by the K-means clustering on the group averaged ERPs corresponding to the other-dominant (15-30% self-voice) and self-dominant (70-85%) voice morphs and to the two types of sound conduction (bone, air). The gray dashed boxes indicate the three time-windows considered for the back-fitting procedure. **B)** Topographic maps associated to each segment. Only Map 4 – indicated by the corresponding color-coded square – was affected by the experimental manipulation.

2.2.4.2.2 Analysis of map parameters

From all seven topographic maps, only Map 4 (Figure 17B) indicated significant interactions involving the fixed effect Parameter, both with the effect of Conduction (F(13, 828.88)=2.41, p=0.003) and the effect of Dominant Voice (F(13, 829.36)=2.58, p=0.002). Moreover, we observed a significant interaction between the effects of Conduction and Dominant Voice (F(1, 836.74)=13.2, p<0.001). To further investigate the nature of these interactions, we ran a separate linear mixed-effects regression for each Parameter of Map 4.

From all parameters, mixed-effects regression revealed significant effects only for the parameter Occurrence. There were both the main effect of Dominant Voice (F(1, 51)=7.19, p=0.01), indicating more occurrence of Map 4 for self-dominant morphs (Figure 18A), and the main effect of Conduction (F(1, 51)=7.01, p=0.011), indicating higher occurrence during air conduction (Figure 18B).

2.2.4.2.3 EEG-Behavior analysis

Following the observed effects on Map 4 Occurrence, we further related Map 4 Occurrence with behavioral performance by conducting linear mixed-effects regressions that contained additional effects of Accuracy and Response Time. We observed the main

effect of Accuracy (F(1, 54.56)=6.01, p=0.017), indicating a negative relationship with overall task accuracy and map occurrence. There was a three-way interaction between the effects of Dominant Voice, Conduction and Accuracy (F(1, 44.7)=4.76, p=0.034). To investigate the nature of this interaction, we performed a separate analysis for each type of sound conduction. Whereas there were no significant effects in bone conduction (all p>0.05), in the trials with air conduction there was a significant interaction between Dominant Voice and Accuracy (F(2, 40.11)=4.39, p=0.019), indicating a stronger negative relationship between Accuracy and Occurrence for self-dominant voice (Figure 18C).

Mixed-effects regression with Response Times as additional factor similarly revealed the main effect of Response Times (F(1, 44.7)=4.76, p=0.034), indicating an overall increase in occurrence with the increase in response times. There was a significant interaction between Response Times and Dominant Voice (F(1, 36.11)=12.26, p=0.001), indicating a stronger positive relationship between response times and Map 4 occurrence for self-dominant, compared to other dominant morphs. Response Times did not interact with Conduction (F(1, 27.32)=2.09, p=0.159), nor was there a significant three-way interaction between Response Times, Conduction and Response Times (F(1, 36.14)=1.69, p=0.202). The effects relating response times to Map 4 occurrence are indicated at Figure 18D.
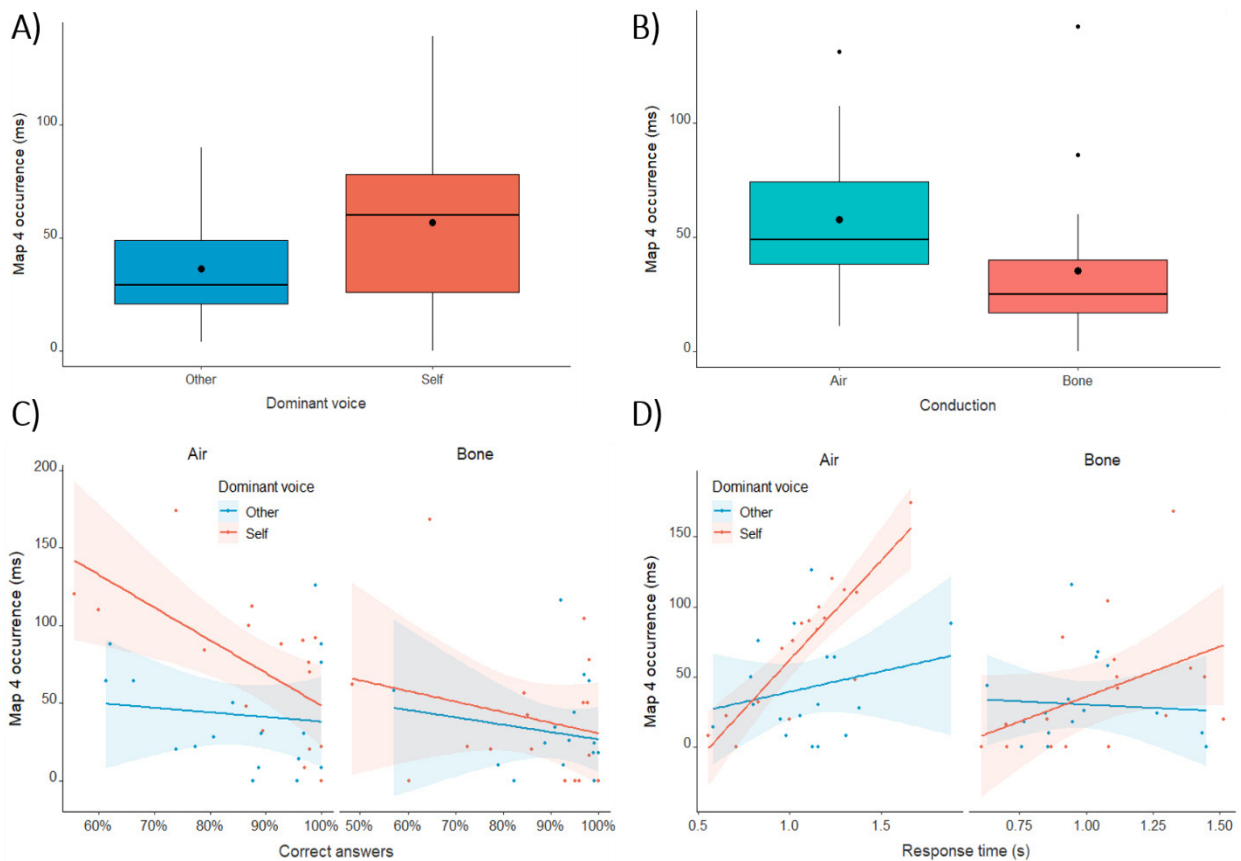
Figure 18. Experimental effects on Map 4 occurrence.
Map 4 occurred more for self-dominant morphs **(A)** and when stimuli were presented through bone conduction **(B)**. Horizontal lines in boxplots indicate median, whereas dots mean values. Map 4 occurrence was negatively correlated to task accuracy **(C)** and positively to response times **(D)**, specifically for self-dominant stimuli presented through air conduction. Shaded areas around linear regressions represent 95% confidence intervals.

2.2.4.2.4 Source localization

The source localization (on the MNI template) associated to Map 4 revealed a network lateralized to the right hemisphere, having the maximum of activation in the right (mainly posterior) insula and concomitant activation in the cingulum and right limbic and para-limbic structures (hippocampus, parahippocampus; Figure 19).
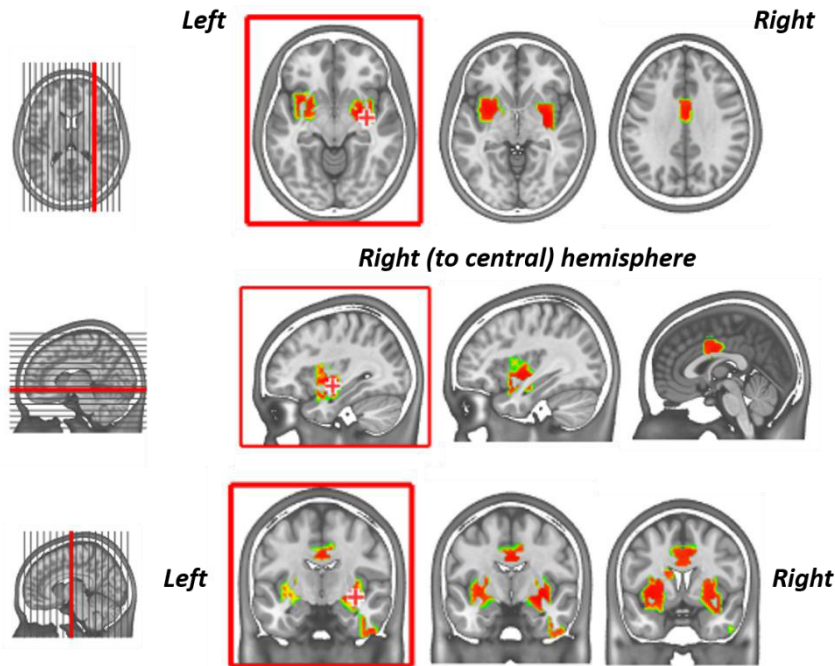
Figure 19. Localization of Map 4.
The brain network associated to self-voice includes the bilateral insulae and putamen with maximal activation occurring on the right hemisphere (red crosses on the brain images in the red boxes). The network also includes the middle cingulum and part of the right inferior temporal pole.

2.2.5   Discussion

By measuring EEG activity during a sensitive SOVD task with stimuli presented through air and bone conduction, we identified a topographic map first occurring around 345 milliseconds after voice onset that was activated more during self-dominant voice morphs and with air conduction. Moreover, the occurrence of this map negatively correlated with accuracy in self-voice trials, which was more prominent during air conduction. Similarly, we observed a positive correlation between map occurrence and response times in self-voice trials, which was, again, stronger when self-voice stimuli were presented through air conduction. Source localization of this map identified a bilateral, but right-dominant network with the maximum activation in posterior insula and further involving putamen, hippocampus, parahippocampal and cingulate gyri.

We propose that this map reflects an auditory self-referencing mechanism that compares the heard voice with the internal self-voice representation. The map occurred more during self-dominant voice morphs, which indicates its specificity for self-voice. Crucially, however, it was related to behavioral performance – only in self-dominant morphs. Specifically, the better participants were in labelling the ambiguous morph as self-dominant and the faster they responded, the less they activated the map. Seen from a different perspective, the worse participants were at recognizing their voice, the more they recruited this map. This is reminiscent of a scenario in which a given resource is used only when needed – e.g. the heavier an object is, the more muscle strength we need in order to lift it. Another analogy could be made with an ability to do a difficult arithmetic task – the better we are at calculus, the faster we would complete the task, thus the less we would recruit arithmetic-associated regions. Accordingly, we suggest that this map represents a neural resource that is used in scenarios of vocal ambiguity, thereby exploited proportionally to the uncertainty associated to self-voice recognition.

This could further be confirmed in the observation that the same map is activated longer with air conduction. Namely, as we have previously shown (Orepic et al., in preparation) and replicated here, participants exhibit a poorer performance at the SOVD task with air, compared to bone conduction. It is thus possible that, since the task is more difficult in this scenario, the map is recruited more. However, both in our previous work and here, the bone-conduction advantage was behaviorally more prominent for other-

dominant morphs. It seems as if bone conduction facilitates discarding an ambiguous voice for being 'not self' rather than labeling it as it as belonging to oneself. One could, therefore, expect that sound conduction would impose a difference in neural activation for other-dominant, and not for self-dominant morphs. Previously, we argued (Orepic et al., in preparation) that the key component leading to the behavioral advantage of bone conduction in SOVD tasks is the multimodal presentation of self-voice stimuli – i.e. besides audition, it involves vestibular and somatosensory processes – which is the case also for natural speech (Emami et al., 2012; Stenfelt, 2011; Tremblay et al., 2003). A differential bone-conduction effect on self-voice perception might, thus, be expected rather in earlier maps and reflected in lower-level, vibrotactile sensory processes. It is possible that such a difference at the neural level is subtle and that our study was underpowered and thus unable to detect it.

Identified source activations further support our proposal. First, right-hemisphere dominance for self-recognition has been consistently reported in various paradigms (Decety & Sommerville, 2003; Feinberg & Keenan, 2005; Frassinetti, Maini, Romualdi, Galante, & Avanzi, 2008; Uddin et al., 2005). Second, insula, where we observed maximal activation, is considered a hub for multisensory integration of exteroceptive and interoceptive signals that serves as a basis for maintaining a coherent representation of our bodily self (Babo-Rebelo, Wolpert, Adam, Hasboun, & Tallon-Baudry, 2016; Craig, 2009; Ionta, Martuzzi, Salomon, & Blanke, 2014; Park et al., 2018; Park & Blanke, 2019; Seth, 2013). Insula activation specific to self-dominant morphs further supports our proposal that self-voice is essentially a multimodal construct (Orepic et al., in preparation). Moreover, (Nakamura et al., 2001) observed activations in parainsular cortex specific to self-voice stimuli and (Shergill et al., 2001) associated insula with inner speech generation. Third, cingulate gyrus has been associated to self-referential processing across different functional domains (for a review see (Northoff et al., 2006)) and specific to self-voice in the study of (Allen et al., 2005). Finally, hippocampus and parahippocampal gyri are well-known to play a role in autobiographical memory retrieval (Cabeza & St Jacques, 2007; Greenberg et al., 2005) as well as in self-referential processing (Kurczek et al., 2015), thus their recruitment in SOVD task might reflect the retrieval of the internal self-voice representation, that is compared to the voice morph.

Contrary to our expectations, we did not observe activity in inferior frontal gyri (IFG), that has been reported in previous neuroimaging self-voice studies (Allen et al., 2005; Kaplan et al., 2008; Nakamura et al., 2001). The main difference in our paradigm compared to those studies is that we employed voice morphing and participants never heard an unmorphed self-voice. It is possible that IFG activation reflects the mere passive response to hearing our voice and since in our study all stimuli contained self-voice features, right IFG might have been activated for both self- and other-dominant voice morphs, thus not surviving the self-other voice contrast. More than just the passive response to self-voice stimuli, the network we identified seems to be related to the cognitive process of relating self-voice stimuli to the internal self-voice representation. Right IFG has been proposed to be involved in processing of self-related stimuli across multiple sensory modalities (Kaplan et al., 2008). It is possible that the self-voice processing in the observed network precedes the activation in the right IFG, by first comparing the heard voice to the internal self-voice representation and relaying the processing to right IFG, where the actual self-attribution might occur. Further studies should delineate the neural mechanisms underlying self-referencing and self-attribution. However, it should be noted that we observed no differences in power-related parameters (e.g. maximal GFP), but in a parameter associated with time, to which fMRI measures are insensitive, which makes it difficult to directly compare these findings to the ones from fMRI studies.

The impact of this work is threefold and encompasses scientific, methodological as well as clinical aspects. First, to the best of our knowledge, we are the first to correlate self-voice-related neural activity with behavioral task performance. This work advances the understanding of the self-voice phenomenon as it provides an exhaustive characterization of spatiotemporal activity related to SOVD, thereby accentuating the importance of its temporal aspects. Second, we further consolidate the importance of multisensory self-voice presentation by demonstrating a reduced processing of the map associated to SOVD task performance with bone- compared to air-conducted self-voices. Finally, this work has a vast clinical importance as it sheds new light on the very mechanism believed to account for auditory-verbal hallucinations – a major source of distress in mental disorders (Harkavy-Friedman et al., 2003) whose underlying principles are still unknown (Wilkinson & Alderson-Day, 2016).

2.2.6 References

Allen, P. P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M., Johns, L. C., & McGuire, P. K. (2005). Neural correlates of the misattribution of self-generated speech. Human Brain Mapping, 26(1), 44–53. https://doi.org/10.1002/hbm.20120

Alzueta, E., Melcón, M., Poch, C., & Capilla, A. (2019). Is your own face more than a highly familiar face? Biological Psychology, 142(January), 100–107. https://doi.org/10.1016/j.biopsycho.2019.01.018

Babo-Rebelo, M., Wolpert, N., Adam, C., Hasboun, D., & Tallon-Baudry, C. (2016). Is the cardiac monitoring function related to the self in both the default network and right anterior insula? Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1708). https://doi.org/10.1098/rstb.2016.0004

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software. https://doi.org/10.18637/jss.v067.i01

Békésy, G. V. (1949). The Structure of the Middle Ear and the Hearing of One's Own Voice by Bone Conduction. Journal of the Acoustical Society of America, 21(3), 217–232. https://doi.org/10.1121/1.1906501

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. Trends in Cognitive Sciences, 8(3), 129–135. https://doi.org/10.1016/j.tics.2004.01.008

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. Neural Computation, 7(6), 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129

Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? Neuroreport, 11(11), R11-6. https://doi.org/10.1586/14737175.7.10.1337

Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. Neuroscience and Biobehavioral Reviews, 47, 717–734. https://doi.org/10.1016/j.neubiorev.2014.10.022

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. Nature Reviews Neuroscience, 13(8), 556–571. https://doi.org/10.1038/nrn3292

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness. Neuron, 88(1), 145–166. https://doi.org/10.1016/j.neuron.2015.09.029

Braun, N., Debener, S., Spychala, N., Bongartz, E., Sörös, P., Müller, H. H. O., & Philipsen, A. (2018). The senses of agency and ownership: A review. Frontiers in Psychology, 9(535), 1–17. https://doi.org/10.3389/fpsyg.2018.00535

Bréchet, L., Brunet, D., Birot, G., Gruetter, R., Michel, C. M., & Jorge, J. (2019). Capturing the spatiotemporal dynamics of self-generated, task-initiated thoughts with EEG and fMRI. NeuroImage, 194, 82–92. https://doi.org/10.1016/j.neuroimage.2019.03.029

Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: CARTOOL. Computational Intelligence and Neuroscience. https://doi.org/10.1155/2011/813870

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. Trends in Cognitive Sciences, 11(5), 219–227. https://doi.org/10.1016/j.tics.2007.02.005

Cabral, J., Kringelbach, M. L., & Deco, G. (2017). Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms. NeuroImage, 160, 84–96. https://doi.org/10.1016/j.neuroimage.2017.03.045

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2016). The effects of stimulus complexity on the preattentive processing of self-generated and nonself voices: An ERP study. Cognitive, Affective and Behavioral Neuroscience, 16(1), 106–123. https://doi.org/10.3758/s13415-015-0376-1

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2018). Stimulus complexity matters when you hear your own voice: Attention effects on self-generated voice processing. International Journal of Psychophysiology, 133(June), 66–78. https://doi.org/10.1016/j.ijpsycho.2018.08.007

Craig, A. D. (2009). How do you feel - now? The anterior insula and human awareness. Nature Reviews Neuroscience, 10(1). https://doi.org/10.1038/nrn2555

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: A social cognitive neuroscience view. Trends in Cognitive Sciences, 7(12), 527–533. https://doi.org/10.1016/j.tics.2003.10.004

Emami, S. F., Pourbakht, A., Sheykholeslami, K., Kamali, M., Behnoud, F., & Daneshi, A. (2012). Vestibular Hearing and Speech Processing. ISRN Otolaryngology, 2012, 1–7. https://doi.org/10.5402/2012/850629

Feinberg, T. E., & Keenan, J. P. (2005). Where in the brain is the self? Consciousness and Cognition, 14(4), 661–678. https://doi.org/10.1016/j.concog.2005.01.002

Ford, J. M., & Mathalon, D. H. (2005). Corollary discharge dysfunction in schizophrenia: Can it explain auditory hallucinations? International Journal of Psychophysiology, 58(2-3 SPEC. ISS.), 179–189. https://doi.org/10.1016/j.ijpsycho.2005.01.014

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nature Reviews Neuroscience, 8(9), 700–711. https://doi.org/10.1038/nrn2201

Frassinetti, F., Maini, M., Romualdi, S., Galante, E., & Avanzi, S. (2008). Is it Mine? Hemispheric Asymmetries in Corporeal Self-recognition. Journal of Cognitive Neuroscience, 20(8), 1507–1516. https://doi.org/10.1162/jocn.2008.20067

Frith, C. D. (1992). The cognitive neuropsychology ofschizophrenia. Erlbaum. https://doi.org/10.1016/0028-3932(93)90120-O

Frith, Christopher D., & Done, D. J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. Psychological Medicine, 19(2), 359-363. https://doi.org/10.4324/9781315630502

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. Psychological Bulletin, 131(1), 76–97. https://doi.org/10.1037/0033-2909.131.1.76

Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., & Bruneau, N. (2015). Is my voice just a familiar voice? An electrophysiological study. Social Cognitive and Affective Neuroscience, 10(1), 101–105. https://doi.org/10.1093/scan/nsu031

Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., Camus, V., & Bruneau, N. (2013). My voice or yours? An electrophysiological study. Brain Topography, 26(1), 72–82. https://doi.org/10.1007/s10548-012-0233-2

Greenberg, D. L., Rice, H. J., Cooper, J. J., Cabeza, R., Rubin, D. C., & LaBar, K. S. (2005). Co-activation of the amygdala, hippocampus and inferior frontal gyrus during autobiographical memory retrieval. Neuropsychologia, 43(5), 659–674. https://doi.org/10.1016/j.neuropsychologia.2004.09.002

Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. Journal of Personality and Social Psychology, 37(2), 147–169. https://doi.org/10.1037/0022-3514.37.2.147

Harkavy-Friedman, J. M., Kimhy, D., Nelson, E. A., Venarde, D. F., Malaspina, D., & Mann, J. J. (2003). Suicide attempts in schizophrenia: the role of command auditory hallucinations for suicide. The Journal of Clinical Psychiatry., 64(8), 871–874.

Hughes, S. M., & Nicholson, S. E. (2010). The processing of auditory and visual recognition of self-stimuli. Consciousness and Cognition, 19(4), 1124–1134. https://doi.org/10.1016/j.concog.2010.03.001

Ionta, S., Martuzzi, R., Salomon, R., & Blanke, O. (2014). The brain network reflecting bodily self-consciousness: A functional connectivity study. Social Cognitive and Affective Neuroscience, 9(12), 1904–1913. https://doi.org/10.1093/scan/nst185

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. Proceedings of the National Academy of Sciences, 106(4), 1245–1248. https://doi.org/10.1073/pnas.0810063106

Kannape, O. A., & Blanke, O. (2012). Agency, gait and self-consciousness. International Journal of Psychophysiology, 83(2), 191–199. https://doi.org/10.1016/j.ijpsycho.2011.12.006

Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q., & Iacoboni, M. (2008). The self across the senses: an fMRI study of self-face and self-voice recognition. Social Cognitive and Affective Neuroscience, 3(3), 218–223. https://doi.org/10.1093/scan/nsn014

Kawahara, H., Morise, M., Banno, H., & Skuk, V. G. (2013). Temporally variable multi-aspect N-way morphing based on interference-free speech representations. In 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2013 (pp. 1–10). IEEE. https://doi.org/10.1109/APSIPA.2013.6694355

Kircher, T., & David, A. S. (2003). Self-consciousness: an integrative approach from philosophy, psychopathology and the neurosciences. In The Self in Neuroscience and Psychiatry. Cambridge University Press. https://doi.org/10.1017/cbo9780511543708.023

Kleiner, M., Brainard, D. H., Pelli, D. G., Ingling, A., Murray, R., & Broussard, A Ingling, R Murray, C. (2007). What's new in Psychtoolbox-3? Perception, 36(14), 1–16. https://doi.org/10.1068/v070821

Kurczek, J., Wechsler, E., Ahuja, S., Jensen, U., Cohen, N. J., Tranel, D., & Duff, M. (2015). Differential contributions of hippocampus and medial prefrontal cortex to self-projection and self-referential processing. Neuropsychologia, 73, 116–126. https://doi.org/10.1016/j.neuropsychologia.2015.05.002

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. Electroencephalography and Clinical Neurophysiology, 48(6), 609–621. https://doi.org/10.1016/0013-4694(80)90419-8

Liu, L., Li, W., Li, J., Lou, L., & Chen, J. (2019). Temporal features of psychological and physical self-representation: An ERP study. Frontiers in Psychology, 10(APR). https://doi.org/10.3389/fpsyg.2019.00785

Lüdecke, D. (2018). sjPlot: Data Visualization for Statistics in Social Science. R package version 2.6.2. https://doi.org/10.5281/zenodo.1308157

Michel, C. M., & Brunet, D. (2019). EEG source imaging: A practical review of the analysis steps. Frontiers in Neurology, 10, 325. https://doi.org/10.3389/fneur.2019.00325

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: A step-by-step tutorial review. Brain Topography, 20(4), 249–264. https://doi.org/10.1007/s10548-008-0054-5

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., … Kojima, S. (2001). Neural substrates for recognition of familiar voices: A PET study. Neuropsychologia, 39(10), 1047–1054. https://doi.org/http://dx.doi.org/10.1016/S0028-3932%2801%2900037-9

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain- A meta-analysis of imaging studies on the self. NeuroImage, 31(1), 440–457. https://doi.org/10.1016/j.neuroimage.2005.12.002

Orepic, P., Kannape, O. A., Faivre, N., & Blanke, O. (2020). Following the inner voice: Approximating internal self-voice representation with bone conduction. In Preparation.

Orepic, Pavo, Rognini, G., Kannape, O. A., Faivre, N., & Blanke, O. (2020). Sensorimotor conflicts induce somatic passivity and loud-en quiet voices in healthy listeners. BioRxiv. https://doi.org/https://doi.org/10.1101/2020.03.26.005843

Park, H. D., Bernasconi, F., Salomon, R., Tallon-Baudry, C., Spinelli, L., Seeck, M., … Blanke, O. (2018). Neural Sources and Underlying Mechanisms of Neural Responses to Heartbeats, and their Role in Bodily Self-consciousness: An Intracranial EEG Study. Cerebral Cortex, 28(7), 2351–2364. https://doi.org/10.1093/cercor/bhx136

Park, H. D., & Blanke, O. (2019). Coupling Inner and Outer Body for Self-Consciousness. Trends in Cognitive Sciences, 23(5), 377–388. https://doi.org/10.1016/j.tics.2019.02.002

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. Electroencephalography and Clinical Neurophysiology, 72(2), 184–187. https://doi.org/10.1016/0013-4694(89)90180-6

R Core Team. (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Reinfeldt, S., Östli, P., Håkansson, B., & Stenfelt, S. (2010). Hearing one's own voice during phoneme vocalization—Transmission by air and bone conduction. The Journal of the Acoustical Society of America, 128(2), 751–762. https://doi.org/10.1121/1.3458855

Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P., & Belin, P. (2008). Investigations of hemispheric specialization of self-voice recognition. Brain and Cognition, 68(2), 204–214. https://doi.org/10.1016/j.bandc.2008.04.007

Sackeim, H. A., & Gur, R. C. (1978). Self-Deception, Self-Confrontation, and Consciousness. In Consciousness and Self-Regulation (pp. 139–197). Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-2571-0_4

Schuerman, W. L., Meyer, A., & McQueen, J. M. (2015). Do we perceive others better than ourselves? A perceptual benefit for noise-vocoded speech produced by an average speaker. PLoS ONE, 10(7), 1–18. https://doi.org/10.1371/journal.pone.0129731

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. Trends in Cognitive Sciences, 17(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Murray, R. M., & McGuire, P. K. (2001). A functional study of auditory verbal imagery. Psychological Medicine, 31(2), 241–253. https://doi.org/10.1017/S003329170100335X

Shergill, Sukhwinder S., White, T. P., Joyce, D. W., Bays, P. M., Wolpert, D. M., & Frith, C. D. (2014). Functional magnetic resonance imaging of impaired sensory prediction in schizophrenia. JAMA Psychiatry, 71(1), 28–35. https://doi.org/10.1001/jamapsychiatry.2013.2974

Shuster, L. I. (1998). The perception of correctly and incorrectly produced /r/. Journal of Speech, Language, and Hearing Research, 41(4), 941–950.

Stenfelt, S. (2011). Acoustic and physiologic aspects of bone conduction hearing. Implantable Bone Conduction Hearing Aids, (71), 10–21. https://doi.org/10.1159/000323574

Stenfelt, S. (2016). Model predictions for bone conduction perception in the human. Hearing Research, 340, 135–143. https://doi.org/10.1016/j.heares.2015.10.014

Todd, N. P. M. A., Cody, F. W. J., & Banks, J. R. (2000). A saccular origin of frequency tuning in myogenic vestibular evoked potentials?: implications for human responses to loud sounds. Hearing Research, 141(1–2), 180–188. https://doi.org/10.1016/S0378-5955(99)00222-1

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. Nature, 423(6942), 866–869. https://doi.org/10.1038/nature01710

Tsakiris, M. (2017). The multisensory basis of the self: From body to identity to others. The Quarterly Journal of Experimental Psychology, 70(4), 597–609. https://doi.org/10.1080/17470218.2016.1181768

Tsakiris, M., & Haggard, P. (2005). Experimenting with the acting self. Cognitive Neuropsychology, 22(3–4), 387–407. https://doi.org/10.1080/02643290442000158

Uddin, L. Q., Kaplan, J. T., Molnar-Szakacs, I., Zaidel, E., & Iacoboni, M. (2005). Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: An event-related fMRI study. NeuroImage, 25(3), 926–935. https://doi.org/10.1016/j.neuroimage.2004.12.018

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wilkinson, S., & Alderson-Day, B. (2016). Voices and Thoughts in Psychosis: An Introduction. Review of Philosophy and Psychology, 7(3), 529–540. https://doi.org/10.1007/s13164-015-0288-6

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. Trends in Cognitive Sciences, 17(6), 263–271. https://doi.org/10.1016/j.tics.2013.04.004

Zarate, J. M., Tian, X., Woods, K. J. P., & Poeppel, D. (2015). Multiple levels of linguistic and paralinguistic features contribute to voice recognition. Scientific Reports, 5, 11475. https://doi.org/10.1038/srep11475

PART II: Robotically-induced self-voice misperceptions

3.1 Study 3: Sensorimotor conflicts induce somatic passivity and louden quiet voices in healthy listeners

Pavo Orepic[1], Giulio Rognini[1], Oliver Alan Kannape[1], Nathan Faivre[2,*], Olaf Blanke[1, 3, *]

1 Laboratory of Cognitive Neuroscience, Center for Neuroprosthetics and Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Switzerland

2 Laboratoire de Psychologie et Neurocognition (LPNC), CNRS UMR 5105, Université Grenoble Alpes, France

3 Faculty of Medicine, University of Geneva, Geneva, Switzerland

* equal contribution

Corresponding author

Olaf Blanke

Bertarelli Chair in Cognitive Neuroprosthetics, Center for Neuroprosthetics & Brain Mind Institute, School of Life Sciences, Campus Biotech, Swiss Federal Institute of Technology (EPFL), 1012 Geneva, Switzerland

E-mail: olaf.blanke@epfl.ch

Tel: +41 (0)21 693 69 21

Fax: +41 (0)21 693 69 22

3.1.1 Abstract

Sensorimotor conflicts are known to alter the perception of accompanying sensory signals and deficits in sensory atten-uation have been observed in schizophrenia. In the auditory domain, self-generated tones or voices (compared to tones or voices presented passively or with sensorimotor delays) have been associated with changes in loudness perception and attenuated neural responses. It has been argued that for sensory signals to be attenuated, predicted and sensory consequences must have a con-sistent spatiotemporal relationship, between button presses and reafferent tactile signals, via predictive sensory signaling, a pro-cess altered in schizophrenia. Here, we investigated auditory sensory attenuation for a series of morphed voices while healthy participants applied sensorimotor stimulations that had no spatiotemporal relationship to the voice stimuli and that have been shown to induce mild psychosis-like phenomena. In two independent groups of participants, we report a loudening of silent voices and found this effect only during maximal sensorimotor conflicts (versus several control conditions). Importantly, conflicting sen-sorimotor stimulation also induced a mild psychosis-like state in the form of somatic passivity and participants who experienced stronger passivity lacked the sensorimotor loudening effect. We argue that this conflict-related sensorimotor loudness amplifica-tion may represent a reduction of auditory self-attenuation that is lacking in participants experiencing a concomitant mild psycho-sis-like state. We interpret our results within the framework of the comparator model of sensorimotor control, and discuss the implications of our findings regarding passivity experiences and hallucinations in schizophrenia.

Keywords

Sensorimotor processing, auditory verbal hallucinations, self-attenuation, passivity experiences, auditory perception, predictive mechanisms

3.1.2 Introduction

Our capacity to process motor signals, their reafferent sensory consequences, and sensory prediction signals is crucial for motor control and perception (Jeannerod, 2006) and for updating internal models of the world (Schultz and Dickinson, 2000). Usually, motor and reafferent signals share similar features in the spatial and temporal domains and according to the comparator model (S J Blakemore et al., 2000; Miall and Wolpert, 1996), movements are accompanied by prediction signals (of their sensory consequences), which are compared with the actual sensory feedback in a feed-forward manner. Under such conditions, spatio-temporal congruence between predicted and reafferent sensory signals is generally associated with self-attribution of the action (Braun et al., 2018; Gallagher, 2000) and the sense of agency: the feeling of being in control of one's movement (Gallagher, 2000; Moore and Fletcher, 2012). A wealth of data has shown that incongruences or sensorimotor conflicts between predicted and reaf-ferent sensory signals lead to the loss of agency and control (David et al., 2008; Farrer et al., 2008; Haggard et al., 2002; MacDonald and Paus, 2003; Sato and Yasuda, 2005; Stetson et al., 2006; Tsakiris et al., 2005).

Sensorimotor conflicts are also known to alter the perception of accompanying sensory signals. Processing of self-generated stimuli is known to be attenuated and proposed to result from a prediction-based cancelation of reafferent sensory signals (Bays et al., 2008; S J Blakemore et al., 2000; Wolpert and Flanagan, 2001). A well-known example is the sensory attenuation of self-generated touch: touches produced by oneself are perceived as weaker compared to externally produced ones, even if applied with the same intensity (Blakemore et al., 1999, 1998; Shergill et al., 2003). Moreover, sensorimotor conflicts accompanying self-generated touches can abolish self-attenuation and thus alter the associated tactile perceptions (S J Blakemore et al., 2000; Kilteni and Ehrsson, 2017a; Weiskrantz et al., 1971).

Perceptual alterations caused by sensorimotor conflicts of upper-limb movements have also been observed in sensory domains other than somatosensation. For instance, studies reported a change in loudness perception of self-generated tones (by a button press), compared to tones presented passively (Sato, 2008; Stenner et al., 2014; Weiss et al., 2011a, 2011b), which was associated with attenuated neural responses (Bansal et al., 2018; Lange, 2011; Martikainen et al., 2005; Mifsud et al., 2016; Schafer and Mar-cus, 1973). Recent studies have demonstrated that such auditory-motor self-attenuation effects can also be obtained for more complex sounds, such as voices (Knolle et al., 2019; Pinheiro et al., 2018). Together, these studies show that motor activity (e.g. a button press) causally associated with the auditory feedback (e.g. a beep or the sound of one's voice) can cause perceptual altera-tions of the latter through a manipulation of its spatiotemporal contingencies.

In general, most of the previous work on sensory alterations based on sensorimotor processes has focused on the investigation of sensory cues for upper-limb actions (e.g. pressing a button). However, the concept of agency, sensorimotor processes and the comparator model have also been applied to movements of the body as a whole (e.g. gait; Kannape and Blanke, 2013, 2012; Menzer et al., 2010), thus affecting the full-body sensorimotor system associated with self-consciousness (Blanke and Metzinger, 2009; Park and Blanke, 2019). Extending previous robotic designs (Blakemore et al., 1999; Shergill et al., 2003; Weiskrantz et al., 1971), Hara et al. (Hara et al., 2011) associated upper-limb sensory prediction signals with reafferent sensory signals at the participants' torso in order to alter the representation of this global, torso-centered bodily system. Using this robotic device, Blanke et al. (Blanke et al., 2014) were able to induce in healthy volunteers systematic changes in illusory own body perceptions (i.e. self-touch) and mild psychosis-like phenomena that depended on sensorimotor conflicts. Specifically, while applying conflicting sensorimotor stimulation between upper-limb movements and tactile feedback on the back participants reported stronger somatic passivity (i.e. that tactile sensations are being imposed upon their body by somebody else) and felt being in a presence of a non-existing alien entity, phenomenologically resembling passivity experiences (Frith et al., 2000; Sass and Parnas, 2003, 2001) and presence hallucinations (Alderson-Day and Fernyhough, 2016; Critchley, 1955; Jaspers, 1990) observed in schizophrenia.

Here, we investigated whether such robotically-mediated sensorimotor conflicts that are able to induce a mild psychosis-like state (Blanke et al., 2014) can also alter voice perception. Alterations of voice perception are highly prevalent in schizophrenia in the form of auditory verbal hallucinations (AVH) – i.e. hearing voices in the absence of a speaker. Given the importance of the comparator model both for somatic passivity and AVH (Ford et al., 2007, 2001; Swiney and Sousa, 2014), we wanted to explore whether robotically-mediated sensorimotor conflicts in healthy participants induce changes in voice perception, resembling the auditory alterations and experiences observed in patients with schizophrenia – specifically loudness alterations (Griffith et al., 1995; Juckel et al., 2008, 2003) and self-other vocal confusion (Frith, 1987; Plaze et al., 2015; Stephane et al., 2018). In two independent experiments, participants were asked to perform repeated upper-limb movements (Blanke et al., 2014), which were conveyed as tactile feedback on their back by the robotic system (Hara et al., 2011). Participants applied sensorimotor stimulation either in a synchronous manner or with a delay while they also performed either the loudness or the self-other voice discrimination task.

3.1.3   Methods

3.1.3.1   Participants

Each of the two separate experiments involved 30 healthy participants from the general population. In experiment 1, 9 participants were male (mean age ±SD: 21.8±2.4 years) and in experiment 2 14 participants were male (23.7±2.4 years). All participants were right-handed according to Edinburgh Handedness Inventory, fluent in French, and without any hearing deficits. Before participating in the experiment, they were screened for eligibility criteria by means of an anamnestic interview investigating medication and substance use, as well as a personal and family history of psychiatric or neurological disorders. Participants were naive to the purpose of the study, gave informed consent in accordance with institutional guidelines (Research project approved by the Comité Cantonal d'Ethique de la Recherche of Geneva) and the Declaration of Helsinki, and received monetary compensation (CHF 20/h).

3.1.3.2   Procedure and materials

We conducted two experiments with the same general procedure and experimental design. Experiment 1 consisted of two and Experiment 2 of three sessions. For the first session of both experiments, participants came with an acquaintance, who also participated in the study, and their voices were recorded. For the second and third sessions (auditory tasks), participants came individually.

3.1.3.3   Auditory tasks

Participants were recorded saying 10 words in French (see supplementary material). Audacity software was used to filter out the background noise and to normalize the recordings for average intensity (-12 dBFS) and duration (500 milliseconds). The pre-processed voice recordings were then entered into TANDEM-STRAIGHT (Kawahara et al., 2013) to generate voice morphs between two participants (e.g. a voice morph could contain 40% of person A's, 60% of person B's voice). Finally, copies of the voice morphs with different sound intensities were created and the resulting audio files were played to participants through a JBL Control 1 Pro speaker placed 1 meter behind them.

During both auditory tasks (loudness, self-other), blindfolded participants repeatedly heard the same word twice, separated by 500 milliseconds. In the loudness task, both words contained the same ratio of the two voices (50% of both participants), but differed in sound intensity. In the self-other task, both words were equally loud, but contained a different ratio of the two voices. In the loudness task, participants reported which of the words they perceived as louder and in the self-other task which of the two words sounded more like their own voice.

Unbeknown to the participants, the first word in each word-pair always sounded the same (50% self-voice, -12 dBFS). The second word varied, either in sound intensity (for the loudness task) or in self-voice percentage (for the self-other task). Six sound intensity levels (dBFS: -14, -13, -12.5, -11.5, -11, -10) and six voice ratios (% self-voice: 15, 30, 45, 55, 70, 85) were chosen based on extensive pilot testing.

3.1.3.4 Robotic system

The robotic system consisted of two integrated units: the front part – a commercial haptic interface (Phantom Omni, SensAble Technologies) – and the back part – a three degree-of-freedom robot (Hara et al., 2011). Participants were seated between the front and back robot and were asked to perform repeated poking movements with their right index finger using the front robot, which was replicated by the movements of the back robot, which applied corresponding touches on their back. This was done either in synchronous (without delay) or asynchronous (with 500 milliseconds delay) fashion, creating different degrees of sensorimotor conflict between the upper limb movement and somatosensory feedback on the back (Blanke et al., 2014).

Experiment 1 and consisted of synchronous and asynchronous sensorimotor conditions. Experiment 2 contained two additional conditions. In the motor-baseline condition, participants performed movements on the front unit, but did not receive the corresponding somatosensory feedback by the back unit. In the touch-baseline condition, the experimenter (not the participant) performed the movements on the front unit, but the participant received the corresponding somatosensory feedback by the back unit. These two conditions served as baselines, as there was no sensorimotor coupling.

In experiment 2 we also tested whether torso-centered tactile feedback (i.e. back) was necessary for the present effects (Park and Blanke, 2019). For this, we added two more conditions in which the same setup was used as in the synchronous and asynchronous conditions, except that tactile feedback was not applied to the back but to the left hand of the participants – i.e. the back unit was placed in front of the participants and adjusted to point downwards in the vertical axis in order to touch their left hand.

3.1.3.5 Experimental design

In experiment 1, participants performed two blocks of each auditory task (loudness and self-other) – one block in the synchronous and another block in the asynchronous condition. Each block started with 60 seconds of robot manipulation, without auditory stimulation, after which an auditory cue indicated the beginning of the actual auditory task. Throughout the auditory tasks, participants continued moving the robot. Importantly, auditory stimuli and participants' movements were not time-locked. Each block contained 60 trials (10 words, each presented with 6 stimulus intensities) presented in a randomized order. The order of tasks (self-other/loudness) and conditions (synchronous/asynchronous) was counterbalanced across participants. An Inter-trial interval of 1 to 1.5 second (randomly jittered), was added to avoid predictability of the stimuli. (Figure 20). The experimental design was created in MATLAB 2017b with Psychtoolbox library (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

3.1.3.6 Statistical analysis

Data of experiment 1 were analyzed with mixed-effects logistic regressions with Response as dependent variable and Condition (synchronous, asynchronous) and Stimulus (levels: 1-6), together with their interaction, as fixed effects. The Response-variable indicates whether participants perceived a stimulus as louder (loudness task) or as sounding more like their own voice (self-other task) compared to the reference stimulus. Random effects included a by-subject random intercept. By-subject random slopes for the main effects were added following model selection based on maximum likelihood. Trials with reaction times greater or smaller than two interquartile ranges from the median for each subject were considered as outliers and excluded.

Analysis for experiment 2 followed a similar approach (two logistic mixed-effects models with Response as a dependent variable). The first model was designed to assess the joint effects of synchrony and location of sensorimotor conflicts, including Condition (synchronous, asynchronous), Location (torso, hand) and Stimulus (levels: 1-6) with interaction terms as fixed effects. The second model extended the first one by investigating the effects of the sensorimotor coupling, regardless of the location. Therefore, it included no main effect of Location and the main effect of Condition had three instead of two levels (synchronous, asynchronous,
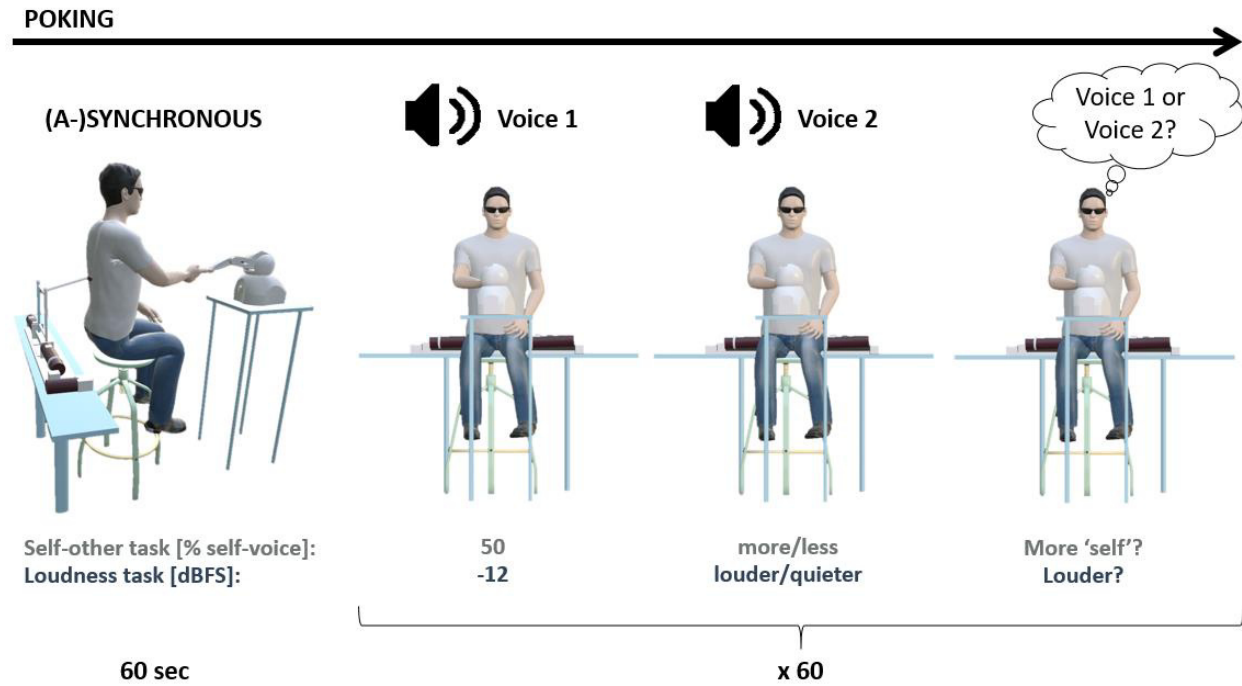
Figure 20. Experimental block design.

baseline). For both experiments, a linear mixed-effects regression was also performed with Reaction Times as a dependent variable. Analysis showed no significant differences between experimental conditions (supplementary material). Questionnaire ratings were assessed by a mixed-effects linear regression and analyzed jointly for experiment 1 and 2, to increase statistical power. As fixed effects, we entered Condition (synchronous, asynchronous) and Question (q1 – q9) with interaction term into the model. As random effects, we had by-subject random intercepts. For the questionnaire items, which significantly differed between the two conditions, we conducted an additional mixed-effects linear regression investigating the fixed effect Location (torso, hand). All analyses were performed with R (R Core Team, 2017), using notably the afex (Singmann et al., 2019), ggplot2 (Wickham, 2016), sjplot (Lüdecke, 2018), lme4 (Bates et al., 2015), and lmerTest (Kuznetsova et al., 2018) packages.

3.1.4   Results

3.1.4.1   Auditory task

3.1.4.1.1   Experiment 1 (Loudness, Self-other)

A mixed-effects logistic regression on loudness judgment revealed higher intercepts in the asynchronous compared to the synchronous condition (estimate=-0.39, Z=-2.14, p=0.03). The model had a main effect of Stimulus (estimate=0.59, Z=9.50, p<0.001) and showed no interaction between the Condition and Stimulus (estimate=0.08, Z=0.05, p=0.12). To further investigate the Stimulus effect observed in the loudness task, we performed the same mixed effects logistic regression for each Stimulus level. Results showed that voices were perceived significantly louder in the asynchronous condition only for the lowest sound intensity level (estimate=-0.5, Z=-2.49, p=0.01) (Figure 21, left), whereas all other stimulus levels did not differ between conditions (supplementary material).

Concerning the self-other discrimination task (Figure 21, right), a mixed-effects logistic regression indicated a main effect of Stimulus (estimate=-2.36, Z=-6.46, p<0.001). Intercepts of the synchronous and asynchronous conditions did not differ in the self-other task (estimate=-0.07, Z=-0.36, p=0.72), nor was there a significant interaction between the Condition and Stimulus (estimate=0.02, Z=0.36, p=0.72).
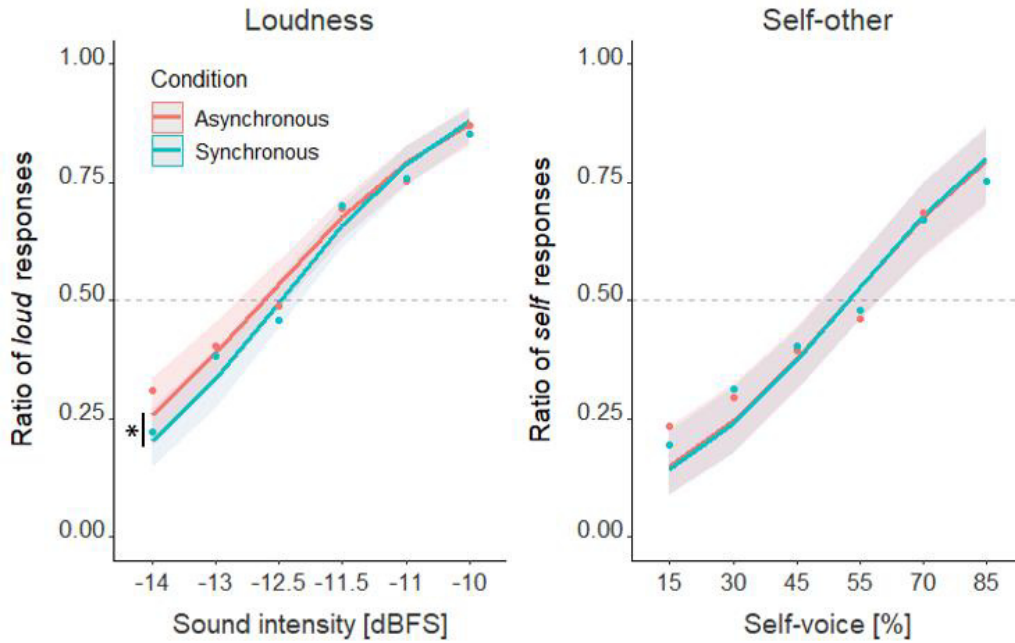
Figure 21. Psychometric curves fitted for the two auditory tasks of experiment 1.
Psychometric curves fitted for the two auditory tasks of experiment 1. The points indicate the rate at which the corresponding voice was perceived as louder (Loudness task) or more resembling own voice (Self-other task) than the baseline. The shaded areas around each curve represent the 95% confidence intervals. Intercept was significantly higher in the asynchronous condition and for the loudness task only, indicating that the quieter voices were perceived as louder. *: p<0.05.

3.1.4.1.2   Experiment 2 (Loudness, hand vs torso)

Experiment 2 replicated the loudness effect observed in experiment 1. In the model assessing both the synchrony and location of sensorimotor conflicts, the intercepts were again significantly higher in the asynchronous compared to synchronous condition (estimate=-0.49, Z=-2.92, p<0.01). The responses differed across stimuli (estimate=0.36, Z=11.22, p<0.001), but there was no significant effect of Location (hand vs. torso) (estimate=-0.3, Z=-1.65, p=0.1). We observed a significant interaction only between the effects of Condition and Stimulus (estimate=0.12, Z=2.51, p=0.01). Interactions between Condition and Location (estimate=0.37, Z=1.54, p=0.12), Stimulus and Location (estimate=0.03, Z=0.69, p=0.49) and a three-way interaction between Condition, Location and Stimulus (estimate=-0.07, Z=-1.11, p=0.27) were not significant.
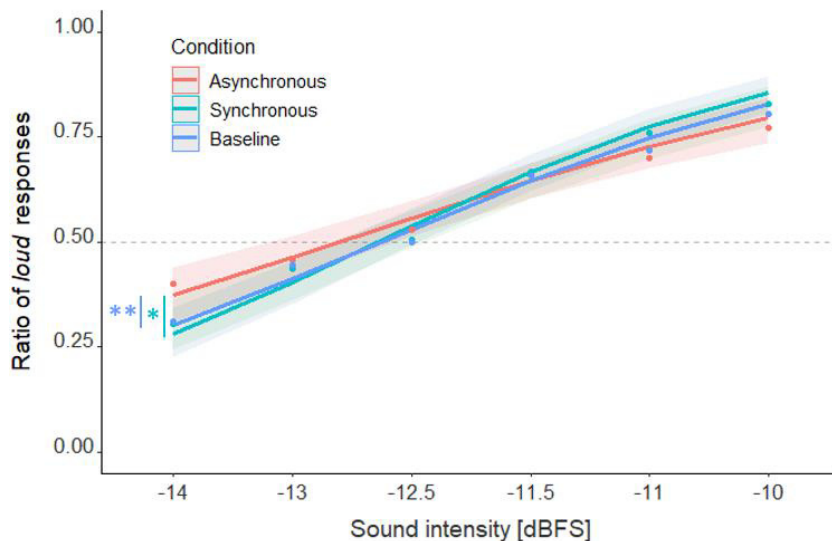
Figure 22. Psychometric curves for Experiment 2.
In experiment 2, intercept in the asynchronous condition was significantly higher than in the synchronous and the baseline conditions, whereas there was no difference between the synchronous and the baseline conditions. *: p<0.05, **: p<0.01.

Analogously to experiment 1, we performed the same mixed effects logistic regression for each Stimulus level, confirming that the difference in loudness perception between the conditions occurred only for the lowest sound intensity level (estimate=-0.35, Z=-2.66, p<0.01, for other levels see supplementary material).

We next addressed the effects of the sensorimotor stimulation, regardless of feedback location. In this model the intercept in the asynchronous condition was higher than the synchronous (estimate=-0.29, Z=-2.23, p=0.03) and the baseline (estimate=-0.51, Z=-3.34, p<0.001), whereas there was no difference between the synchronous and baseline conditions (estimate=-0.17, Z=-1.29, p=0.2) (Figure 22).

3.1.4.2   Subjective experience

The linear mixed-model analysis revealed that participants experienced stronger somatic passivity in the asynchronous versus synchronous condition (Figure 23A) (estimate=-0.83, t(66.94)=-2.88, p<0.01) and rated illusory self-touch significantly stronger in the synchronous versus asynchronous condition (Figure 23B) (estimate=0.64, t(67.54)=2.56, p=0.01), without any significant difference between conditions in other questionnaire items (all p>0.05).

For the two significant questionnaire items, an additional mixed-effects linear regression was applied, in order to investigate the effects of Location, showing that somatic passivity was significantly stronger when sensorimotor conflicts were applied on the torso vs. hand (Figure 23C) (estimate=1.34, t(88.56)=3.08, p<0.01). Self-touch ratings did not differ between the two locations (Figure 23D) (estimate=-0.1, t(87.66)=-0.24, p=0.81).

To assess the relationship between subjective experience and auditory perception, we ran mixed-effects logistic regression with significant questionnaire items (Passivity and Self-touch) as additional factors and divided participants in two groups – those with a positive asynchronous-synchronous rating difference (Passivity+, Self-touch+) and those with a negative or zero difference (Passivity-, Self-touch-). Model showed a significant interaction between Passivity and Condition (estimate=0.39, Z=2.04, p=0.04) (supplementary material). Investigation of the interaction showed that loudness perception was altered only in Passivity- group (Figure 24,
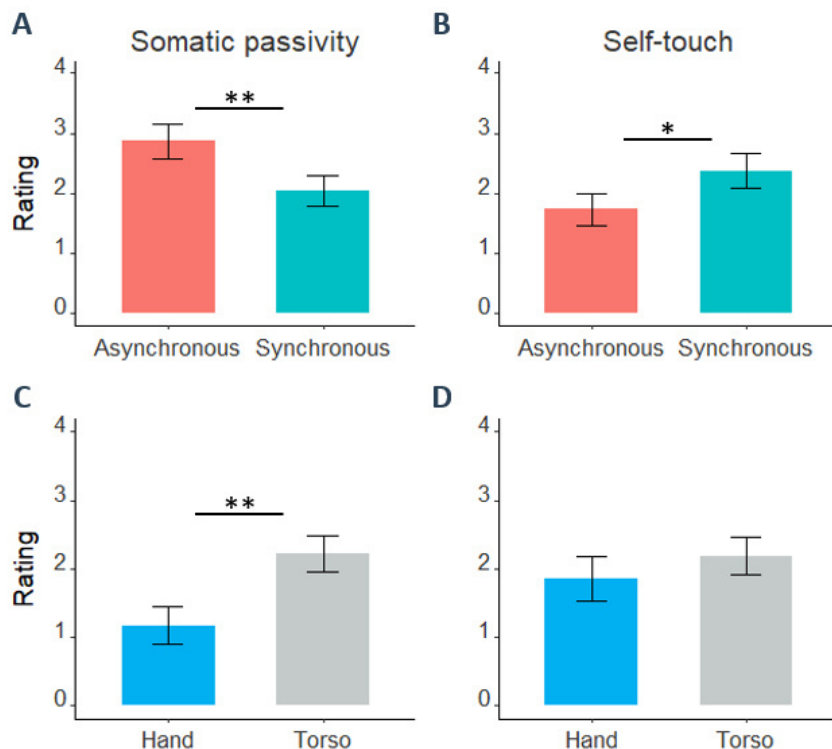
Figure 23. Subjective experience.
Abscissa of every bar plot indicates either the two experimental conditions (A, B: Synchronous, Asynchronous) or the location of sensorimotor conflicts (C, D: Hand, Torso) and ordinate the corresponding Likert-scale ratings. Height of a bar plot indicates the mean rating and error bars its standard error. Somatic passivity sensations were reported significantly higher in the asynchronous condition (**A**) and with sensorimotor conflicts applied to torso compared to hand (**C**). Self-touch impressions were stronger in the synchronous condition (**B**) but equally strong for both locations (**D**). *: p<0.05, **: p<0.01

left) (Condition: estimate=-0.54, Z=-3.71, p<0.001; Stimulus: estimate=0.47, Z=7.07, p<0.001; Condition-Stimulus interaction: estimate=0.12, Z=3.05, p<0.01), with no difference between conditions in Passivity+ group (Figure 24, right) (supplementary material). There were no significant interactions between Self-touch and Condition (supplementary material).
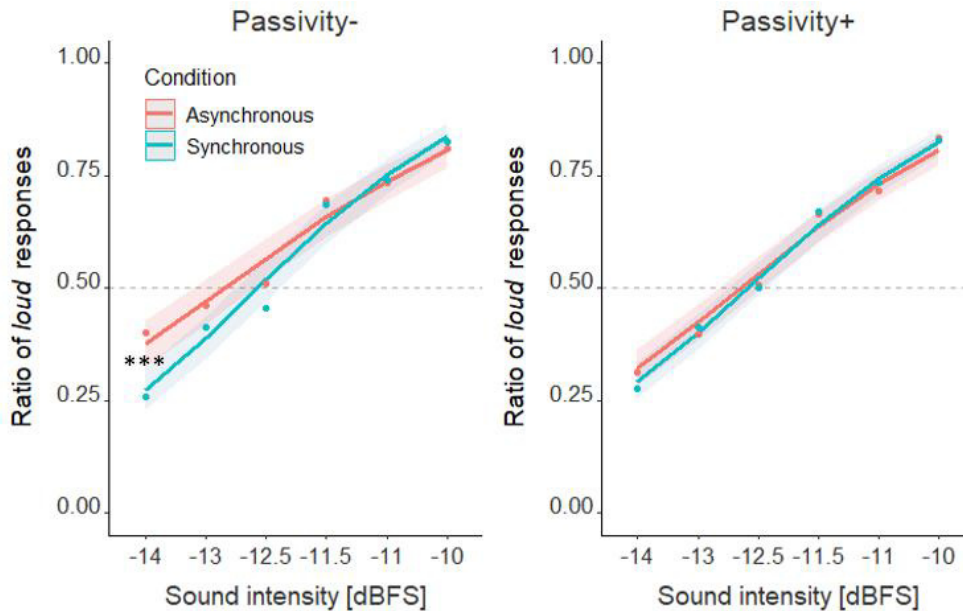
Figure 24. Loudness effect and somatic passivity.
Loudness effect and somatic passivity. Quiet voices were amplified only for the participants not experiencing somatic passivity during the experiment (Passivity-, left). With somatic passivity (Passivity+, right) there was no change in voice perception. ***: p<0.001

3.1.5   Discussion

Replicating the induction of somatic passivity based on sensorimotor stimulation in a healthy population using a robotic procedure (Blanke et al., 2014; Hara et al., 2014; Salomon et al., 2020) we investigated potential links with voice perception and clinical phenomenology (i.e. AVHs) and demonstrate that voice perception is modulated by sensorimotor stimulation with somatosensory feedback. We confirmed this somatosensory-motor effect on auditory perception in two independent cohorts in two studies. Specifically, quiet voices were perceived as louder in the asynchronous condition, differing from voices heard in synchronous and baseline conditions.

Changes in perception during actions are usually interpreted within the comparator model framework: self-generated movements are accompanied by sensory predictions, which cause an attenuation of the reafferent sensory signals, especially if they are received in spatiotemporal congruency (S J Blakemore et al., 2000; Miall and Wolpert, 1996). Thus, in order for the sensory signal to be attenuated, predicted and reafferent sensory consequences must have a consistent spatiotemporal relationship such as pushing a response button with one's right index finger attenuating processing of tactile stimuli at the fingertip, via predictive sensory signaling (S J Blakemore et al., 2000; Shergill et al., 2003). Related work has extended these findings to auditory perception, showing that auditory processing of a sound triggered by a button press is attenuated (Knolle et al., 2019; Martikainen et al., 2005). Lack of predictive mechanisms is associated with decreases in sensory attenuation and perceived as amplification of the sensory stimuli accompanying actions (stronger touches (Kilteni and Ehrsson, 2017b; Shergill et al., 2003; Teufel et al., 2010); louder sounds (Sato, 2008; Stenner et al., 2014; Weiss et al., 2011a, 2011b)).

The present findings extend sensory attenuation research in two ways. First, there was no time-locking between our participants' movements and the auditory stimuli they were asked to judge. Participants manipulated the robot independently from the sounds and the auditory task – ruling out the possibility that classical effects linked to the comparator framework and associated with trial-by-trial sensory comparisons between an action and its sensory consequences account for the present effects. Secondly, perceptual

changes in both experiments were only present in the asynchronous condition, accentuating the importance of temporal aspects (between movement and somatosensory feedback) of sensorimotor conflicts. In experiment 1 and 2, we observed a difference in loudness perception between the asynchronous and synchronous conditions and in experiment 2, additionally, observed that perception in the asynchronous condition is the deviating one, as it alone differed from baseline conditions. Crucially, the perception in the spatially-conflicting, yet synchronous condition did not differ from the no-conflict conditions (touch- and motor-baseline), suggesting that mainly the temporal conflict, present only in the asynchronous condition, drives the present perceptual effects. Temporal conflicts have been shown to cause a loss of agency, by manipulating sensory action consequences of upper-limb movements and related losses of hand movement agency (David et al., 2008; Farrer et al., 2008; Haggard et al., 2002; MacDonald and Paus, 2003; Sato and Yasuda, 2005; Stetson et al., 2006; Tsakiris et al., 2005). When extending such manipulations to a torso-centered bodily system (Blanke and Metzinger, 2009; Park and Blanke, 2019), other-agency changes can be introduced (Blanke et al., 2014; Salomon et al., 2020), together with a state of an altered bodily self-consciousness, including the alien agent (Blanke et al., 2014; Salomon et al., 2020). We argue that loudness amplification, observed solely in the asynchronous condition, may represent a reduction of auditory self-attenuation, resulting from such other-agency-related alterations in bodily self-consciousness.

Deficits in self-attenuation have been observed in schizophrenia. When healthy participants overestimate the externally-applied stimulation, arguably due to sensory attenuation for actively produced movements, individuals with schizophrenia perform differently, suggesting an alteration of corrections related to sensory attenuation (S. J. Blakemore et al., 2000; Shergill et al., 2005), compatible with neural responses between self- and externally-generated sounds in individuals with schizophrenia (Ford et al., 2007, 2001) and in healthy individuals depending on hallucination proneness (Asai, 2016; Teufel et al., 2010; Whitford et al., 2017). Our results in healthy subjects support this inverse relationship by demonstrating a lack of loudness increase only in the hallucinating group, extending previous data on changes in self-other voice discrimination in early psychosis patients with passivity symptoms (Salomon et al., 2020). Interestingly, somatic passivity was experienced more strongly when receiving torso-centered bodily feedback, compared to hand feedback, a finding not observed for illusory self-touch. As, in addition, the strength of illusory self-touch did not interact with the loudening effect, we suggest that torso-centered manipulations involving sensations related to another agent (passivity experience) interfere more strongly with voice perception than more focal somatosensory feedback (hand). Collectively, these findings suggest that asynchronous torso-centered sensorimotor stimulation (1) induces a mild psychosis-like state in the form of somatic passivity and (2) is associated with a loudening of voices, however, that (3) experiencing somatic passivity leads to a lack of voice loudening, suggesting a reduction in self-attenuation mechanisms.

Differences in divided attention between asynchronous vs. synchronous conditions cannot account for these effects, because (1) both sensorimotor conditions contained a strong conflict and both induced an altered mental state (asynchronous: somatic passivity; synchronous: self-touch), because (2) reaction times revealed no differences between both sensorimotor conditions, and because (3) the effect was only observed in one auditory task. Although it is further known that auditory perception is altered during movement (Reznik and Mukamel, 2019), movements in the synchronous and motor-only conditions were not accompanied by changes in auditory perception, suggesting the necessity of a temporal conflict for the present loudness effect. The present sensorimotor conflicts did not affect self-other voice discriminability. It is possible that a motor component involving speech production is necessary to observe a misattribution of one's own voice in healthy individuals, as is argued to occur in AVHs (Frith, 1987; Moseley et al., 2013; Nazimek et al., 2012; Stephan et al., 2009). The orthogonal sensorimotor stimulation, as tested in the present experiments, changes loudness, but not identity of the heard voice.

To the best of our knowledge, this is the first study to demonstrate that temporal sensorimotor conflicts in the somatosensory domain can affect voice perception even if the auditory stimulus is not systematically linked to the movement. We found that healthy listeners heard quiet voices as louder when exposed to asynchronous sensorimotor stimulation related to somatic passivity experiences. We argue that this amplification represents a reduction in self-attenuation mechanisms, reminiscent of altered voice perception in psychiatric populations. Together, our findings extend the understanding of subjective and perceptual alterations caused by conflicting sensorimotor processing and suggest that passivity experiences and voice perception rely, at least partly, on common sensorimotor brain mechanisms.

3.1.6   References

Alderson-Day, B., Fernyhough, C., 2016. Auditory verbal hallucinations: Social, but how? J. Conscious. Stud. 23, 163–194. https://doi.org/10.1007/978-1-4614-0959-5_9

Asai, T., 2016. Self is "other", other is "self"_ poor self-other discriminability explains schizotypal twisted agency judgment. Psychiatry Res. 246, 593–600. https://doi.org/10.1016/j.psychres.2016.10.082

Bansal, S., Ford, J.M., Spering, M., 2018. The function and failure of sensory predictions. Ann. N. Y. Acad. Sci. 1426, 199–220. https://doi.org/10.1111/nyas.13686

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Bays, P.M., Wolpert, D.M., Haggard, P., Rossetti, Y., Kawato, M., 2008. Predictive attenuation in the perception of touch, in: Sensorimotor Foundations of Higher Cognition. Oxford University Press, USA, pp. 339–358.

Blakemore, S.J., Frith, C.D., Wolpert, D.M., 1999. Spatio-temporal prediction modulates the perception of self-produced stimuli. J. Cogn. Neurosci. 11, 551–559. https://doi.org/10.1162/089892999563607

Blakemore, S. J., Smith, J., Steel, R., Johnstone, E.C., Frith, C.D., 2000. The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. Psychol. Med. 30, 1131–1139. https://doi.org/10.1017/S0033291799002676

Blakemore, S J, Wolpert, D., Frith, C., 2000. Why can't you tickle yourself? Neuroreport 11, R11-6. https://doi.org/10.1586/14737175.7.10.1337

Blakemore, S.J., Wolpert, D.M., Frith, C.D., 1998. Central cancellation of self-produced tickle sensation. Nat. Neurosci. 1, 635–640. https://doi.org/10.1038/2870

Blanke, O., Metzinger, T., 2009. Full-body illusions and minimal phenomenal selfhood. Trends Cogn. Sci. 13, 7–13. https://doi.org/10.1016/j.tics.2008.10.003

Blanke, O., Pozeg, P., Hara, M., Heydrich, L., Serino, A., Yamamoto, A., Higuchi, T., Salomon, R., Seeck, M., Landis, T., Arzy, S., Herbelin, B., Bleuler, H., Rognini, G., 2014. Neurological and robot-controlled induction of an apparition. Curr. Biol. 24, 2681–2686. https://doi.org/10.1016/j.cub.2014.09.049

Brainard, D.H., 1997. The Psychophysics Toolbox. Spat. Vis. 10, 433–436. https://doi.org/10.1163/156856897X00357

Braun, N., Debener, S., Spychala, N., Bongartz, E., Sörös, P., Müller, H.H.O., Philipsen, A., 2018. The senses of agency and ownership: A review. Front. Psychol. 9, 1–17. https://doi.org/10.3389/fpsyg.2018.00535

Critchley, M., 1955. The idea of a presence. Acta Psychiatr. Scand. 30, 155–168. https://doi.org/10.1111/j.1600-0447.1955.tb06055.x

David, N., Newen, A., Vogeley, K., 2008. The "sense of agency" and its underlying cognitive and neural mechanisms. Conscious. Cogn. 17, 523–534. https://doi.org/10.1016/j.concog.2008.03.004

Farrer, C., Bouchereau, M., Jeannerod, M., Franck, N., 2008. Effect of Distorted Visual Feedback on the Sense of Agency. Behav. Neurol. 19, 53–57. https://doi.org/10.1155/2008/425267

Ford, J.M., Gray, M., Faustman, W.O., Roach, B.J., Mathalon, D.H., 2007. Dissecting corollary discharge dysfunction in schizophrenia. Psychophysiology 44, 522–529. https://doi.org/10.1111/j.1469-8986.2007.00533.x

Ford, J.M., Mathalon, D.H., Heinks, T., Kalba, S., Faustman, W.O., Roth, W.T., 2001. Neurophysiological Evidence of Corollary Discharge Dysfunction in Schizophrenia. Am J Psychiatry 158, 2069–2071.

Frith, C.D., 1987. The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. Psychol. Med. 17, 631–648. https://doi.org/10.1017/S0033291700025873

Frith, C.D., Blakemore, S.J., Wolpert, D.M., 2000. Abnormalities in the awareness and control of action. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 355, 1771–1788. https://doi.org/10.1098/rstb.2000.0734

Gallagher, S., 2000. Philosophical conceptions of the self: Implications for cognitive science. Trends Cogn. Sci. 4, 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Griffith, J., Hoffer, L.D., Adler, L.E., Zerbe, G.O., Freedman, R., 1995. Effects of sound intensity on a midlatency evoked response to repeated auditory stimuli in schizophrenic and normal subjects. Psychophysiology 32, 460–466.

Haggard, P., Clark, S., Kalogeras, J., 2002. Voluntary action and conscious awareness. Nat. Neurosci. 5, 382–385. https://doi.org/10.1038/nn827

Hara, M., Rognini, G., Evans, N., Blanke, O., Yamamoto, A., Bleuler, H., Higuchi, T., 2011. A novel approach to the manipulation of body-parts ownership using a bilateral master-slave system, in: IEEE International Conference on Intelligent Robots and Systems. pp. 4664–4669. https://doi.org/10.1109/IROS.2011.6048519

Hara, M., Salomon, R., van der Zwaag, W., Kober, T., Rognini, G., Nabae, H., Yamamoto, A., Blanke, O., Higuchi, T., 2014. A novel manipulation method of human body ownership using an fMRI-compatible master-slave system. J. Neurosci. Methods 235, 25–34. https://doi.org/10.1016/j.jneumeth.2014.05.038

Jaspers, K., 1990. Über leibhaftige Bewußtheiten (Bewußtheitstäuschungen), ein psychopathologisches Elementsymptom, in: Gesammelte Schriften Zur Psychopathologie. Springer, Berlin, Heidelberg, pp. 413–420. https://doi.org/10.1007/978-3-642-62027-0_8

Jeannerod, M., 2006. The origin of voluntary action. History of a physiological concept. C. R. Biol. 329, 354–62. https://doi.org/10.1016/j.crvi.2006.03.017

Juckel, G., Gallinat, J., Riedel, M., Sokullu, S., Schulz, C., Möller, H.J., Müller, N., Hegerl, U., 2003. Serotonergic dysfunction in schizophrenia assessed by the loudness dependence measure of primary auditory cortex evoked activity. Schizophr. Res. 64, 115–124. https://doi.org/10.1016/S0920-9964(03)00016-1

Juckel, G., Gudlowski, Y., Müller, D., Özgürdal, S., Brüne, M., Gallinat, J., Frodl, T., Witthaus, H., Uhl, I., Wutzler, A., Pogarell, O., Mulert, C., Hegerl, U., Meisenzahl, E.M., 2008. Loudness dependence of the auditory evoked N1/P2 component as an indicator of serotonergic dysfunction in patients with schizophrenia - A replication study. Psychiatry Res. 158, 79–82. https://doi.org/10.1016/j.psychres.2007.08.013

Kannape, O.A., Blanke, O., 2013. Self in motion: sensorimotor and cognitive mechanisms in gait agency. J. Neurophysiol. 110, 1837–1847. https://doi.org/10.1152/jn.01042.2012

Kannape, O.A., Blanke, O., 2012. Agency, gait and self-consciousness. Int. J. Psychophysiol. 83, 191–199. https://doi.org/10.1016/j.ijpsycho.2011.12.006

Kawahara, H., Morise, M., Banno, H., Skuk, V.G., 2013. Temporally variable multi-aspect N-way morphing based on interference-free speech representations, in: 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2013. IEEE, pp. 1–10. https://doi.org/10.1109/APSIPA.2013.6694355

Kilteni, K., Ehrsson, H.H., 2017a. Body ownership determines the attenuation of self-generated tactile sensations. Proc. Natl. Acad. Sci. 114, 8426–8431. https://doi.org/10.1073/pnas.1703347114

Kilteni, K., Ehrsson, H.H., 2017b. Sensorimotor predictions and tool use: Hand-held tools attenuate self-touch. Cognition 165, 1–9. https://doi.org/10.1016/j.cognition.2017.04.005

Kleiner, M., Brainard, D.H., Pelli, D.G., Ingling, A., Murray, R., Broussard, A Ingling, R Murray, C., 2007. What's new in Psychtoolbox-3? Perception 36, 1–16. https://doi.org/10.1068/v070821

Knolle, F., Schwartze, M., Schröger, E., Kotz, S.A., 2019. Auditory predictions and prediction errors in response to self-initiated vowels. Front. Neurosci. 13, 1146.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2018. lmerTest Package: Tests in Linear Mixed Effects Models. J. Stat. Softw. 82, 1–26. https://doi.org/10.18637/jss.v082.i13

Lange, K., 2011. The reduced N1 to self-generated tones: An effect of temporal predictability? Psychophysiology 48, 1088–1095. https://doi.org/10.1111/j.1469-8986.2010.01174.x

Lüdecke, D., 2018. sjPlot: Data Visualization for Statistics in Social Science. R package version 2.6.2. https://doi.org/10.5281/zenodo.1308157

MacDonald, P.A., Paus, T., 2003. The role of parietal cortex in awareness of self-generated movements: A transcranial magnetic stimulation study. Cereb. Cortex 13, 962–967. https://doi.org/10.1093/cercor/13.9.962

Martikainen, M.H., Kaneko, K.I., Hari, R., 2005. Suppressed responses to self-triggered sounds in the human auditory cortex. Cereb. Cortex 15, 299–302. https://doi.org/10.1093/cercor/bhh131

Menzer, F., Brooks, A., Halje, P., Faller, C., Vetterli, M., Blanke, O., 2010. Feeling in control of your footsteps: Conscious gait monitoring and the auditory consequences of footsteps. Cogn. Neurosci. 1, 184–192. https://doi.org/10.1080/17588921003743581

Miall, R.C., Wolpert, D.M., 1996. Forward models for physiological motor control. Neural networks 9, 1265–1279.

Mifsud, N.G., Oestreich, L.K.L., Jack, B.N., Ford, J.M., Roach, B.J., Mathalon, D.H., Whitford, T.J., 2016. Self-initiated actions result in suppressed auditory but amplified visual evoked components in healthy participants. Psychophysiology. https://doi.org/10.1111/psyp.12605

Moore, J.W., Fletcher, P.C., 2012. Sense of agency in health and disease: A review of cue integration approaches. Conscious. Cogn. 21, 59–68. https://doi.org/10.1016/j.concog.2011.08.010

Moseley, P., Fernyhough, C., Ellison, A., 2013. Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. Neurosci. Biobehav. Rev. 37, 2794–2805. https://doi.org/10.1016/j.neubiorev.2013.10.001

Nazimek, J.M., Hunter, M.D., Woodruff, P.W.R., 2012. Auditory hallucinations: Expectation-perception model. Med. Hypotheses 78, 802–810. https://doi.org/10.1016/j.mehy.2012.03.014

Park, H.D., Blanke, O., 2019. Coupling Inner and Outer Body for Self-Consciousness. Trends Cogn. Sci. 23, 377–388. https://doi.org/10.1016/j.tics.2019.02.002

Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spat. Vis. 10, 437–442. https://doi.org/10.1163/156856897X00366

Pinheiro, A.P., Schwartze, M., Kotz, S.A., 2018. Voice-selective prediction alterations in nonclinical voice hearers. Sci. Rep. 8, 1–10. https://doi.org/10.1038/s41598-018-32614-9

Plaze, M., Mangin, J.F., Paillère-Martinot, M.L., Artiges, E., Olié, J.P., Krebs, M.O., Gaillard, R., Martinot, J.L., Cachia, A., 2015. "Who is talking to me?" - Self-other attribution of auditory hallucinations and sulcation of the right temporoparietal junction. Schizophr. Res. 169, 95–100. https://doi.org/10.1016/j.schres.2015.10.011

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Found. Stat. Comput.

Reznik, D., Mukamel, R., 2019. Motor output, neural states and auditory perception. Neurosci. Biobehav. Rev. 96, 116–126. https://doi.org/10.1016/j.neubiorev.2018.10.021

Salomon, R., Progin, P., Griffa, A., Rognini, G., Do, K.Q., Conus, P., Marchesotti, S., Bernasconi, F., Hagmann, P., Serino, A., Blanke, O., 2020. Sensorimotor Induction of Auditory Misattribution in Early Psychosis. Schizophr. Bull. https://doi.org/10.1093/schbul/sbz136

Sass, L.A., Parnas, J., 2003. Schizophrenia, consciousness, and the self. Schizophr. Bull. 29, 427–444.

Sass, L.A., Parnas, J., 2001. Phenomenology of Self-Disturbances in Schizophrenia: Some Research Findings and Directions. Philos. Psychiatry, Psychol. 8, 347–356. https://doi.org/10.1353/ppp.2002.0027

Sato, A., 2008. Action observation modulates auditory perception of the consequence of others' actions. Conscious. Cogn. 17, 1219–1227. https://doi.org/10.1016/j.concog.2008.01.003

Sato, A., Yasuda, A., 2005. Illusion of sense of self-agency: Discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. Cognition 94, 241–255. https://doi.org/10.1016/j.cognition.2004.04.003

Schafer, E.W.P., Marcus, M.M., 1973. Self-Stimulation Alters Human Sensory Brain Responses. Science 181, 175–177.

Schultz, W., Dickinson, A., 2000. Neural coding of prediction errors. Annu. Rev. Neurosci. 23, 473–500.

Shergill, S.S., Bays, P.M., Frith, C.D., Wolpert, D.M., 2003. Two Eyes for an Eye: The Neuroscience of Force Escalation. Science 301, 187–187. https://doi.org/10.1126/science.1085327

Shergill, S.S., Samson, G., Bays, P.M., Frith, C.D., Wolpert, D.M., 2005. Evidence for sensory prediction deficits in schizophrenia. Am. J. Psychiatry 162, 2384–2386. https://doi.org/10.1176/appi.ajp.162.12.2384

Singmann, H., Bolker, B., Westfall, J., Aust, F., 2019. afex: Analysis of Factorial Experiments. R package version 0.23-0.

Stenner, M.P., Bauer, M., Sidarus, N., Heinze, H.J., Haggard, P., Dolan, R.J., 2014. Subliminal action priming modulates the perceived intensity of sensory action consequences. Cognition 130, 227–235. https://doi.org/10.1016/j.cognition.2013.11.008

Stephan, K.E., Friston, K.J., Frith, C.D., 2009. Dysconnection in Schizophrenia: From abnormal synaptic plasticity to failures of self-monitoring. Schizophr. Bull. 35, 509–527. https://doi.org/10.1093/schbul/sbn176

Stephane, M., Burton, P., Meriwether, D., Yoon, G., 2018. "Other" tags for "Self"-generated speech in patients with auditory verbal hallucinations, an fMRI study. Schizophr. Res. 202, 410–411. https://doi.org/10.1016/j.schres.2018.06.056

Stetson, C., Cui, X., Montague, P.R., Eagleman, D.M., 2006. Motor-Sensory Recalibration Leads to an Illusory Reversal of Action and Sensation. Neuron 51, 651–659. https://doi.org/10.1016/j.neuron.2006.08.006

Swiney, L., Sousa, P., 2014. A new comparator account of auditory verbal hallucinations: how motor prediction can plausibly contribute to the sense of agency for inner speech. Front. Hum. Neurosci. 8, 675. https://doi.org/10.3389/fnhum.2014.00675

Teufel, C., Kingdon, A., Ingram, J.N., Wolpert, D.M., Fletcher, P.C., 2010. Deficits in sensory prediction are related to delusional ideation in healthy individuals. Neuropsychologia 48, 4169–4172. https://doi.org/10.1016/j.neuropsychologia.2010.10.024

Tsakiris, M., Haggard, P., Franck, N., Mainy, N., Sirigu, A., 2005. A specific role for efferent information in self-recognition. Cognition 96, 215–231. https://doi.org/10.1016/j.cognition.2004.08.002

Weiskrantz, L., Elliott, J., Darlington, C., 1971. Preliminary observations on tickling oneself. Nature 230, 598–599. https://doi.org/10.1038/230598a0

Weiss, C., Herwig, A., Schütz-Bosbach, S., 2011a. The self in action effects: Selective attenuation of self-generated sounds. Cognition 121, 207–218. https://doi.org/10.1016/j.cognition.2011.06.011

Weiss, C., Herwig, A., Schütz-Bosbach, S., 2011b. The self in social interactions: Sensory attenuation of auditory action effects is stronger in interactions with others. PLoS One 6, 16–18. https://doi.org/10.1371/journal.pone.0022723

Whitford, T.J., Mitchell, A.M., Mannion, D.J., 2017. The ability to tickle oneself is associated with level of psychometric schizotypy in non-clinical individuals. Conscious. Cogn. 52, 93–103. https://doi.org/10.1016/j.concog.2017.04.017

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wolpert, D.M., Flanagan, J.R., 2001. Motor prediction. Curr. Biol. 11, R729–R732.

3.1.7 Supplementary material

3.1.7.1 Words

Participants were recorded saying 10 words in French (clou, fouet, hache, lame, lutte, os, rat, sang, scie, ver). The words were chosen from the list of 100 negatively-valenced words, as rated by 20 schizophrenic patients and 97 healthy participants (Jalenques et al., 2013).

3.1.7.2 Experiment 1 (Loudness, Self-other)

In the loudness task, participants perceived the target voice as louder than the reference in 57.4% of all trials, with an average of 26.6% for the lowest (-14 dBFS) and 86.2% for the highest stimulus level (-10 dBFS), indicating that we effectively sampled the parameter space of the task (main Figure 2, left). In the self-other task, participants perceived their voice as the dominant one in 46.9% of trials, with an average of 21.4% for the lowest (15% self-voice present) and 75.1% for the highest stimulus level (85% self-voice) (main Figure 2, right).

To further investigate the effect observed in the loudness task, we performed the same mixed effects logistic regression for each Stimulus level, allowing us to identify the sound intensity levels driving the difference in loudness perception. Results showed that voices were perceived significantly louder in the asynchronous condition only for the lowest sound intensity level (quiet sounds; level 1: estimate=-0.5, Z=-2.49, p=0.01), whereas all other stimulus levels did not differ between conditions (2: estimate=-0.11, Z=-0.62, p=0.53; 3: estimate =-0.13, Z =-0.79, p =0.43; 4: estimate=0, Z=0.03, p=0.98; 5: estimate=0.04, Z=0.19, p=0.85; 6: estimate=-0.15, Z=-0.6, p=0.55). Thus, quiet voices were perceived as louder in the asynchronous condition, whereas there were no perceptual differences for louder voices between the two experimental conditions (Figure 21, left).

3.1.7.3 Experiment 2 (Loudness, hand vs torso)

Participants perceived the target voice as louder than the reference in 57.9% of all trials, with an average of 34.9% for the lowest (-14 dBFS) and 79.5% for the highest stimulus level (-10 dBFS) (Figure 22).

Analogously to experiment 1, we performed the same mixed effects logistic regression for each Stimulus level, confirming that the difference in loudness perception between the conditions occurred only for the lowest sound intensity level (quiet sounds; level 1: estimate=-0.35, Z=-2.66, p<0.01). All other stimulus levels did not differ between both sensorimotor conditions (2: estimate=-0.02, Z=-0.18, p=0.86; 3: estimate=-0.06, Z=-0.53, p=0.6; 4: estimate=-0.03, Z=-0.2, p=0.84; 5: estimate=0.11, Z=0.13, p=0.4; 6: estimate=0.14, Z=0.92, p=0.36).

The model investigating the effects of the sensorimotor coupling, regardless of the location, could be designed in one more way, as it is mentioned in the main text. Namely, the main effect of Condition contained four instead of three levels (synchronous, asynchronous, motor-baseline, touch-baseline). In this model the intercept in the asynchronous condition was significantly higher than the ones from all other conditions (synchronous: estimate=-0.31, Z=-2.56, p=0.01; motor-baseline: estimate=-0.4, Z=-2.42, p=0.02; touch-baseline: estimate=-0.59, Z=-3.57, p<0.001). There were no other significant differences between the intercepts of the other three conditions (synchronous and motor-baseline: estimate=-0.09, Z=-0.52, p=0.61; synchronous and touch-baseline: estimate=-0.28, Z=-1.68, p=0.09; motor-baseline and touch-baseline: estimate=-0.2, Z=-0.1, p=0.32). Thus, this model confirms the findings of the model mentioned in the main text.

3.1.7.4 Reaction times

For both experiments, a linear mixed-effects regression was also performed with Reaction Times as a dependent variable. The model contained the same fixed and random effects as the model with Response as a dependent variable, except that a polynomial expansion of Stimulus variable to the power of two was added, to account for the non nonlinear distribution of reaction times around the point of subjective equivalence (i.e., reaction times became shorter with more extreme stimulus levels).

Participants responded to the auditory stimuli on average in 1.23 seconds in the loudness task and in the self-other task in 1.42 seconds. In experiment 1, there were no significant differences in reaction times between the conditions in any of the tasks (loudness: estimate = 0, t(30) = 0.26, p = 0.8; self-other: estimate = 0, t(29.7) = -0.12, p = 0.9). There was a main effect of Stimulus in both tasks (loudness: estimate = -3.21, t(3462) = -5.86, p < 0.001; self-other: estimate = -3.36, t(3414) = -5.41, p < 0.001), without significantly interacting with the conditions (loudness: estimate = 0.96, t(3462) = 1.29, p = 0.2; self-other: estimate = -1.34, t(3414) = -1.53, p = 0.13).

In experiment 2, there were no significant effects of Condition (estimate = -0.02, t(46) = -0.63, p = 0.53) nor Location (estimate = 0.05, t(36) = 1.29, p = 0.21) on reaction times. The model showed a main effect of Stimulus (estimate = -2.27, t(6818) = -2.48, p = 0.01), but no significant interactions between the fixed effects (Condition and Location: estimate = -0.02, t(6822) = -0.87, p = 0.39; Condition and Stimulus: estimate = 0.38, t(6818) = 0.29, p = 0.77; Location and Stimulus: estimate = 0.21, t(6818) = 0.17, p = 0.87; Location, Condition and Stimulus: estimate = 0.67, t(6817) = 0.37, p = 0.71). As in experiment 1, reaction times did not differ between conditions.

3.1.7.5 Questionnaire

After each questionnaire block, participants filled in the following 9-item questionnaire (extended from Blanke et al., 2014) by rating how strongly they agreed with each item on a Likert scale from 0 (Not at all) to 6 (Very strong). Participants reported the subjective feeling of touching oneself ("I felt as if I was touching my back by myself."), the strength of somatic passivity ("I felt as if someone else was touching my back.") and presence hallucination ("I felt as if someone was standing close to me."). Additionally, we investigated whether sensorimotor conflicts and voiced stimuli affected each other on a subjective level. On one hand, we examined whether the voiced stimulation imposed an identity to the potentially evoked presence ("I felt as if my friend was standing close to me."; "I felt as if someone else than my friend was standing close to me."). On the other hand, we explored whether sensorimotor conflicts biased the perceived auditory ambiguity ("I felt as if I heard my friend's voice more often than my own voice."; "I felt as if I heard a voice that was neither my friend's nor mine."). Finally, participants were asked to report changes in bodily sensations experienced during the corresponding block ("While hearing the words, I felt changes in my body sensations (e.g. lighter, warmer, I felt tingling sensations etc.)"), and to describe them by means of an open self-report ("If answer is between 1 and 6, please describe the changes in body sensations.").

3.1.7.6 Spontaneous negative reports

As a final part of our questionnaire, there was an open-type question, where participants were asked to describe the bodily sensations they might have experienced during the experiment. Several participants freely reported about emotional changes and 11 participants felt negative emotions and, in addition, reported them only during the asynchronous condition (I felt "annoyed", "anxious" (2), "bad", "choking", "painful", "sad", "stressed" (2), "tense", "worried") (only one instance was reported during the synchronous condition ("bad")). A chi-square test of independence was performed to examine the relation between experimental conditions (synchronous, asynchronous) and emotional valence (negative, non-negative). This test revealed that negative emotions were more likely during the asynchronous condition (X2(2, N=120) = 9.25 p < .01).

Out of 60 participants, 31 (52%) had higher somatic passivity in the asynchronous condition (Passivity+ group, Figure 24, right) and all participants who spontaneously reported negative emotions in the asynchronous condition also belonged to the Passivity+ group. 23 (38%) participants reported higher self-touch sensations in synchronous condition. Thus, the robot-induced sound effect and psychosis-like state were associated with negative emotional valence, as participants spontaneously reported negative emotions (Figure 25), especially during the asynchronous condition and in individuals experiencing somatic passivity (Passivity+). One could argue that the appearance of negative emotions was due to the fact that the participants were hearing negative words, yet all experimental conditions contained the same negative auditory stimuli, and negative sensations were reported predominately after the asynchronous condition. This represents another phenomenological resemblance to clinical voice-hearing, as negative valence has been proposed to be a determining factor for separating clinical from healthy AVHs (Daalman et al., 2011; Lawrence et al., 2010; Woods et al., 2015).
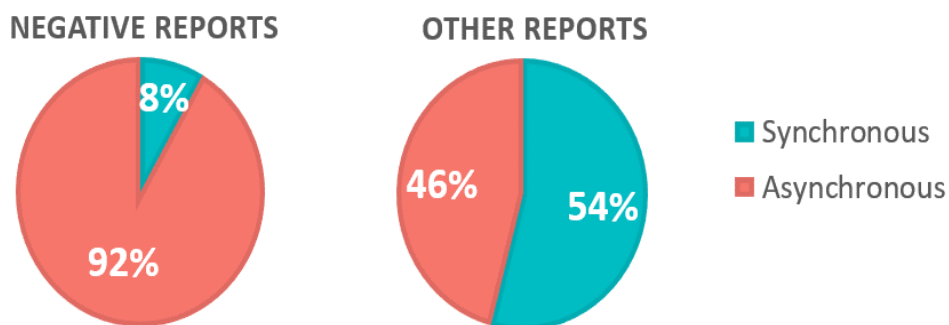
Figure 25. Open self-reports.
When asked to describe changes in bodily sensations – such as feeling warmer or lighter – experienced during the experimental blocks, some participants spontaneously reported feeling negative emotions. Such negative reports were associated with the asynchronous condition.

3.1.7.7 Auditory perception and subjective experience

To assess the relationship between the subjective experience and auditory perception, we ran the same mixed-effects logistic regression with significant questionnaire items (Passivity and Self-touch) as additional factors. For both items, participants were divided in two groups – those with a positive difference in ratings between the asynchronous and synchronous conditions (Passivity+ and Self-touch+) and those with a negative or zero difference (Passivity- and Self-touch-).

The model which had Passivity as an additional factor showed significant effects of Condition (estimate=0.59, Z=-4.27, p<0.001) and Stimulus (estimate=0.39, Z=14.88, p<0.001), with a significant interaction between the two factors (estimate=0.13, Z=3.56, p<0.001). The effect of Passivity borderlined with significance (estimate=-0.28, Z=-1.65, p=0.09) and interacted with the effect of Condition (estimate=0.39, Z=2.04, p=0.04). Passivity did not interact with Stimulus (estimate=0.05, Z=1.37, p=0.17) and the three-way interaction between Condition, Stimulus and Passivity was not significant (estimate=-0.08, Z=-1.54, p=0.12). Investigation of the interaction between Passivity and Condition showed that loudness perception was altered only in Passivity- group (Figure 24, left) (Condition: estimate=-0.54, Z=-3.71, p<0.001; Stimulus: estimate=0.47, Z=7.07, p<0.001; Condition-Stimulus interaction: estimate=0.12, Z=3.05, p<0.01), with no difference between conditions in Passivity+ group (Figure 24, right) (the effect of Condition: estimate=-0.15, Z=-1.05, p=0.29; Stimulus: estimate = 0.51, Z = 9.57, p < 0.001; Condition and Stimulus interaction: estimate = 0.04, Z = 1.17, p = 0.24).

The model with Self-touch as an additional factor also showed significant effects of Condition (estimate=-0.43, Z=-3.52, p<0.001) and Stimulus (estimate=0.41, Z=18.02, p<0.001), also with a significant interaction between the two factors (estimate=0.1, Z=3.01, p<0.01). However, the effect of Self-touch was not significant (estimate=0.12, Z=0.6, p=0.55). It did not interact with the effects of Condition (estimate=0.1, Z=0.48, p=0.63) nor Stimulus (estimate=0.02, Z=0.45, p=0.65), nor was the interaction between Condition, Stimulus and Self-touch significant (estimate=-0.01, Z=-0.26, p=0.8).

Additionally, we ran monotonic (Spearman) correlation analyses between the significant questionnaire and auditory task findings. It indicated a negative monotonic relationship between the effects of synchrony on task performance and the intensity of somatic passivity (Figure 26) (ρ=-0.3, p=0.03). Specifically, the difference between the asynchronous and synchronous conditions in loudness perception of quiet voices (i.e. individual average responses for the lowest stimulus level) negatively correlated with the difference in somatic passivity experienced during the asynchronous and synchronous conditions (i.e. individual questionnaire ratings). The same correlation analysis between the differences in task performance and the intensity self-touch sensation between the asynchronous and synchronous condition did not indicate a significant relationship (ρ=0.2, p=0.14).
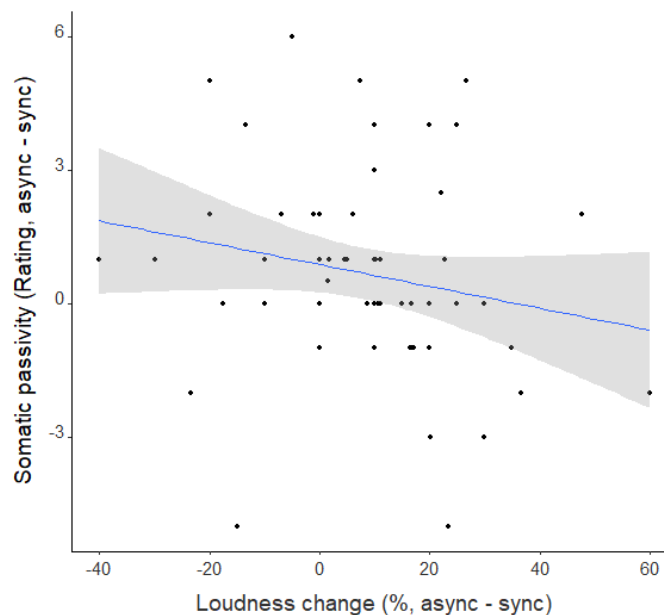
Figure 26. Somatic passivity negatively correlates with voice amplification.
Somatic passivity negatively correlates with voice amplification. Each dot represents an individual increase in loudness perception of quiet voices (abscissa) and somatic passivity (ordinate) between the asynchronous and synchronous conditions. Blue line represents a linear regression describing the negative monotonic relationship between the two, with the shaded area indicating its 95% confidence interval.

3.1.7.8 Supplementary references

Blanke, O., Pozeg, P., Hara, M., Heydrich, L., Serino, A., Yamamoto, A., Higuchi, T., Salomon, R., Seeck, M., Landis, T., Arzy, S., Herbelin, B., Bleuler, H., Rognini, G., 2014. Neurological and robot-controlled induction of an apparition. Curr. Biol. 24, 2681–2686. https://doi.org/10.1016/j.cub.2014.09.049

Daalman, K., Boks, M.P.M., Diederen, K.M.J., De Weijer, A.D., Blom, J.D., Kahn, R.S., Sommer, I.E.C., 2011. The same or different? A phenomenological comparison of auditory verbal hallucinations in healthy and psychotic individuals. J. Clin. Psychiatry 72, 320–325. https://doi.org/10.4088/JCP.09m05797yel

Jalenques, I., Enjolras, J., Izaute, M., 2013. Valence émotionnelle des mots. L'Encéphale 39, 189–197.

Lawrence, C., Jones, J., Cooper, M., 2010. Hearing voices in a non-psychiatric population. Behav. Cogn. Psychother. 38, 363–373. https://doi.org/10.1017/S1352465810000172

Woods, A., Jones, N., Alderson-Day, B., Callard, F., Fernyhough, C., 2015. Experiences of hearing voices: Analysis of a novel phenomenological survey. The Lancet Psychiatry 2, 323–331. https://doi.org/10.1016/S2215-0366(15)00006-1

3.2 Study 4: Breathing affects self-other voice discrimination in a bodily state with increased otherness

Pavo Orepic[1], Hyeong-Dong Park[1], Giulio Rognini[1], Nathan Faivre[2,*], Olaf Blanke[1, 3,*]

1 Laboratory of Cognitive Neuroscience, Center for Neuroprosthetics and Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Switzerland

2 Laboratoire de Psychologie et Neurocognition (LPNC), CNRS UMR 5105, Université Grenoble Alpes, France

3 Faculty of Medicine, University of Geneva, Geneva, Switzerland

* equal contribution

Corresponding author

Olaf Blanke

Bertarelli Chair in Cognitive Neuroprosthetics, Center for Neuroprosthetics & Brain Mind Institute, School of Life Sciences, Campus Biotech, Swiss Federal Institute of Technology (EPFL), 1012 Geneva, Switzerland

E-mail: olaf.blanke@epfl.ch

Tel: +41 (0)21 693 69 21

Fax: +41 (0)21 693 69 22

3.2.1 Abstract

A growing number of studies have focused on identifying cognitive processes that are modulated by interoceptive signals. Here we investigated whether interoception affects self-processing, by assessing changes in self-voice perception as a function of respiratory and cardiac cycles. Considering the fundamental role interoception plays in bodily self-consciousness, we additionally applied conflicting sensorimotor stimulation inducing a state characterized by a loss of self and increased otherness, and investigated its effects in self-other voice perception. Our data reveal that breathing, but not heartbeat, affects self-voice perception, by showing that participants (N = 30) discriminated self-voice from other voices better during inspiration, while being in the state of increased otherness and especially when hearing voices of other people. Loudness judgement of equivalent self-related stimuli was unaffected by breathing. Combining interoception and voice perception with self-monitoring framework, these data extend recent findings on breathing-dependent cognition to self-processing.

Keywords

Breathing, interoception, self-voice, self-other voice discrimination, bodily self-consciousness, self-monitoring, somatic passivity

Statement of Relevance

Respiration and heartbeat signals, traditionally placed under the umbrella of autonomic interoceptive processing, have recently been investigated in cognitive neuroscience and identified as contributors to various cognitive processes. Compared to numerous reports relating cognition to cardiac cycle, breathing-phase dependency has been demonstrated only for a few cognitive functions, and interestingly, revealed a consistent advantage of the inspiration phase. Here, we combined psychophysics with robotics and voice-morphing technology to investigate the impact of respiration and heartbeat cycles on self-voice perception and on bodily self-processing. Our data associate breathing to self-voice perception by demonstrating an inspiration-driven advantage in self-other voice discrimination, which is further affected by experimentally-induced alterations in bodily self-perception. These findings are of broad relevance because they describe breathing-dependent effects in self-processing and thereby intersect three seemingly distinct branches of modern neuroscience – self-monitoring, voice processing and interoception.

3.2.2 Introduction

Interoception refers to the processing of afferent sensory signals originating from the inside of the body, such as signals from the heart, lungs, or intestines. Although traditionally considered as mostly unconscious signals of the autonomic nervous system, recent research has shown that heartbeat and respiration signals can also affect many perceptual and cognitive processes. For instance, there is consistent evidence showing that external sensory stimuli (visual, tactile or auditory) are perceived differently when presented in different phases of the cardiac cycle (Birren, Cardon, & Phillips, 1963; Motyka et al., 2019; Sandman, McCanne, Kaiser, & Diamond, 1977). Cardiac phase dependency has also been observed for pain perception (Wilkinson, McIntyre, & Edwards, 2013) and emotional processing (Garfinkel et al., 2014), as well as cognitive functions such as memory (Pfeifer et al., 2017) or social cognition (Azevedo, Garfinkel, Critchley, & Tsakiris, 2017). By comparison, only few recent studies demonstrated respiratory phase dependency in emotion distinction and memory recall (Zelano et al., 2016), as well as visuospatial perception (Perl et al., 2019), and visual pattern recognition (Nakamura, Fukunaga, & Oku, 2018). Interestingly, all studies investigating the impact of respiration on cognitive functions reported enhanced performance during inspiration as compared to the phase of expiration, arguably due to inspiration-driven neural synchronization of task-relevant cortical and subcortical regions (Heck, Kozma, & Kay, 2019; Perl et al., 2019; Zelano et al., 2016).

Interoceptive signals are also known to be constituting brain mechanisms of bodily self-consciousness (BSC) (Park & Blanke, 2019; Seth & Tsakiris, 2018), as based on multisensory and sensorimotor bodily signals (Blanke, Slater, & Serino, 2015). This is corroborated by experimental evidence demonstrating that the integration of conflicting interoceptive and exteroceptive signals (afferent sensory signals encompassing vision, audition, somatosensation, gustation and olfaction) can lead to altered states of BSC (Adler, Herbelin, Similowski, & Blanke, 2014; Aspell et al., 2013; Suzuki, Garfinkel, Critchley, & Seth, 2013). Similar alterations of BSC have also been reported by using a robotic device (Hara et al., 2011), which creates sensorimotor conflicts between a participant's upper

limb movements and touch sensations on the back. Namely, poking movements performed with the front part of the robotic device (placed in front of participants) are replicated by the back part of the device (Figure 26), resulting in the corresponding tactile stimuli on participants' back (synchronous stimulation). Moreover, adding a temporal delay between the participants' movements and the tactile stimulation delivered on the back (asynchronous stimulation) induces an alteration of BSC characterized by differences in self-location (Blanke et al., 2014) and in self-monitoring (Faivre et al., 2020). Importantly, these and related stimulations also induce the feeling that another person is in the room (Blanke et al., 2014; Salomon et al., 2020), a loss of self-agency (Sato & Yasuda, 2005; Tsakiris, Haggard, Franck, Mainy, & Sirigu, 2005), and somatic passivity (i.e. the impression that someone else is applying tactile sensations on our body), compatible with an altered state of BSC, characterized by misperceiving self as other (Blanke et al., 2014; Salomon et al., 2020).

Compared to its well-established relationship to bodily self-processing, the impact of interoceptive processes on the perception of self-related stimuli (such as perception of one's own voice or face) has only rarely been investigated (Ambrosini, Finotti, Ruben, Tsakiris, & Ferri, 2019). Accordingly, it is not known whether inspiration-enhanced performance extends to the perception of self-related stimuli (such as perception of one's own face or voice) and whether this is further modulated by sensorimotor stimulations modulating BSC (e.g. faster responses in a self-face recognition task occurring only in an altered state of BSC). Here, we investigated cardiac and respiratory phase dependency of self-voice perception. We recorded heartbeat and respiration signals of healthy participants performing two self-related auditory tasks (self-other voice discrimination; loudness judgment) (Orepic, Rognini, Kannape, Faivre, & Blanke, 2020). We investigated whether self-voice perception would differ in trials occurring during different parts of respiratory (inspiration, expiration) and heartbeat (systole, diastole) cycles. Following previously reported breathing effects on cognition (Nakamura et al., 2018; Perl et al., 2019; Zelano et al., 2016) and heartbeat effects on self-face perception (Ambrosini et al., 2019), we predicted better performance in auditory tasks during inspiration and during systole. Additionally, we explored the effects of BSC modulations on respiration and cardiac phase dependency in self-voice perception. Simultaneously with performing the auditory tasks, participants manipulated a robotic device known to induce BSC alterations through sensorimotor stimulation (Hara et al., 2011). Based on our previous findings (Blanke et al., 2014; Salomon et al., 2020), we investigated whether conflicting sensorimotor stimulation able to induce systematic changes in conscious self-other experience would modify self-other voice discrimination selectively dependent on respiratory and cardiac cycles.

3.2.3   Method

3.2.3.1   Participants

The study involved 30 right-handed participants (9 male, mean age ± SD: 21.8 ± 2.4 years old), chosen from the general population, fluent in French and naïve to the purpose of the study. All participants reported no hearing deficits and no history of psychiatric or neurological disorders. Participants gave informed consent in accordance with institutional guidelines (protocol 2015-00092, approved by the Comité Cantonal d'Ethique de la Recherche of Geneva) and the Declaration of Helsinki, and received monetary compensation (CHF 20/h). Two participants were excluded from the analysis due to noisy ECG and respiration signals. The analysis reported here was done on the sample from experiment 1 of our previous study (Orepic et al., 2020).

3.2.3.2   Auditory tasks

Pairs of gender-matched acquaintances participated in this study. Participants' voices were recorded while saying 10 words in French (Zoom H6 Handy recorder; see supplementary material for an overview). Background noise removal and normalization of the recordings for average intensity (-12 dBFS) and duration (500 milliseconds) was done in Audacity software. These recordings were used to generate voice morphs spanning a voice identity continuum between two participants (acquaintances) by using TANDEM-STRAIGHT (Kawahara, Morise, Banno, & Skuk, 2013) (e.g. a voice morph can be generated such that it contains 30% of person A's, 70% of person B's voice). Finally, the generated voice morphs were recreated six times, such that each copy contained different sound intensity. Participants heard the recordings through a speaker (JBL Control 1 Pro) placed 1 meter behind them. The experimental design was created in MATLAB 2017b with Psychtoolbox library (Kleiner et al., 2007).

Participants performed two auditory tasks – self-other task and loudness task. During both tasks, blindfolded participants repeatedly heard the same word twice, while the first word in each word-pair always sounded the same (50% self-voice, -12 dBFS). In the self-other task, the second word was always equally loud as the first word (-12 dBFS), but varied in participants' self-voice percentage (% self-voice: 15, 30, 45, 55, 70, 85). In each trial, participants were instructed to indicate which of the two words sounded more like their own voice by clicking on a button. In the loudness task, the second word always contained the same ratio of the two

voices (50% of both participants), but varied in sound intensity (dBFS: -14, -13, -12.5, -11.5, -11, -10). Accordingly, participants were instructed to choose the louder of the two words. Six sound intensity levels and six voice ratios were chosen based on extensive pilot testing.

3.2.3.3   Robotic system

The robotic system consisted of two integrated units: the front part – a commercial haptic interface (Phantom Omni, SensAble Technologies) – and the back part – a three degree-of-freedom robot (Hara et al., 2011) (Figure 1). Participants were seated between the front and back robot and were asked to perform repeated poking movements with their right index finger using the front robot. Participants' pokes were replicated by the back robot, thus applying corresponding touches on participants' backs. The touches were mediated by the robot either in synchronous (without delay) or asynchronous (with 500 milliseconds delay) fashion, creating different degrees of sensorimotor conflict between the upper limb movement and somatosensory feedback on the back (Blanke et al., 2014; Faivre et al., 2020; Salomon et al., 2020). Participants carried out a familiarization session, after which they were asked to perform poking movements in any direction (touches could be applied on their backs in a region with a 200 mm x 250 mm surface).

3.2.3.4   Experimental design

The study consisted of two experimental sessions. For the first session, participants came with an acquaintance (a friend), who also participated in the study. Both of them were screened for eligibility criteria, after which their voices were recorded. For the second session, each participant came individually and performed the auditory tasks. Respiration and heartbeat signals were recorded throughout the entire second session.

The second session comprised two blocks of each auditory task (loudness and self-other) – one block with the synchronous and another block with the asynchronous stimulation (Figure 27). The order of blocks (loudness synchronous, loudness asynchronous, self-other synchronous, self-other asynchronous) was counterbalanced across participants. Each block started with 60 seconds of robot manipulation, without auditory stimulation, after which an auditory cue indicated the beginning of the actual auditory task. Throughout the auditory tasks, participants continued moving the robot and auditory stimuli were not time-locked to participants' movements. Each block contained 60 randomly ordered trials (10 word pairs, each presented with 6 stimulus intensities). The
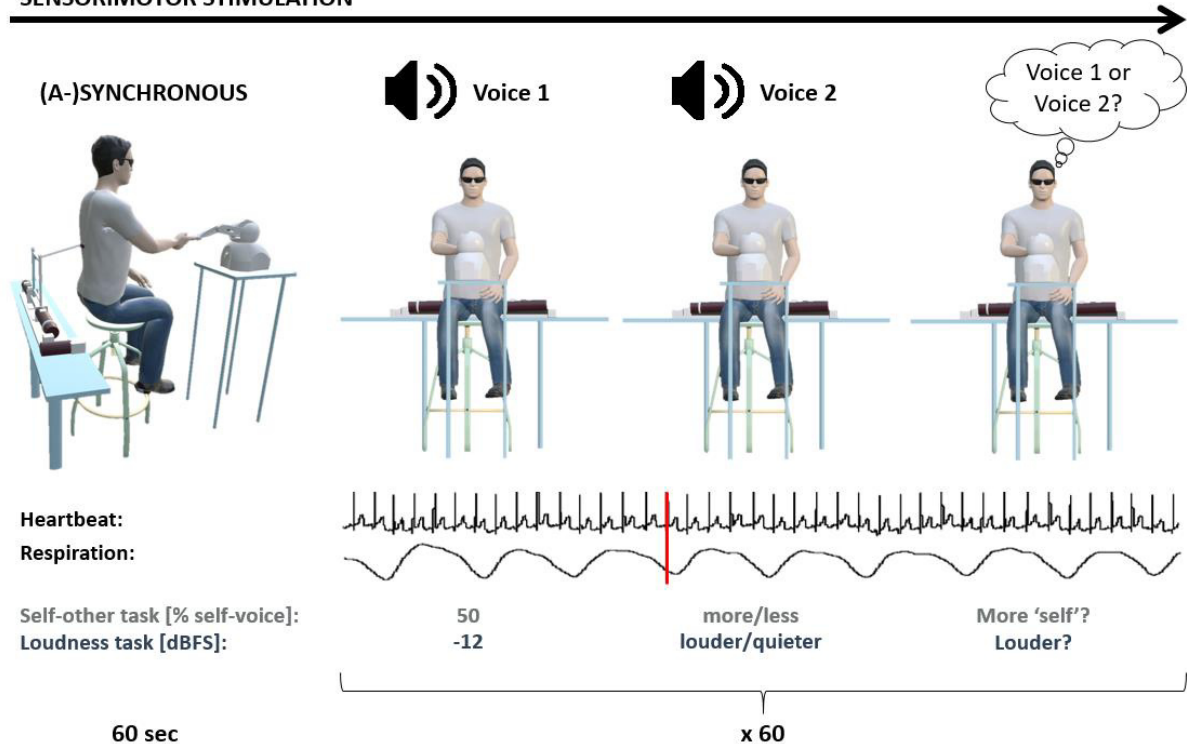

Figure 27. Experimental block design.
Heartbeat and respiration signals were continuously recorded during sensorimotor stimulation and auditory tasks (see main text for description). Cardiac and respiratory phase was extracted at the onset of the second vocal stimulus (red line). Adapted from (Orepic, Rognini, Kannape, Faivre, & Blanke, 2020).

words within a pair were separated by 500 milliseconds and an inter-trial interval of 1 to 1.5 seconds (randomly jittered) was added to avoid predictability of the stimuli.

At the end of the second session, participants performed two additional blocks (synchronous and asynchronous) in which they passively listened to the same voice morphs while manipulating the robot. Instead of performing an auditory task during these blocks, participants were asked to fill out a short questionnaire after each block (Likert scale from 0 (not at all) to 6 (very strong), adapted from Blanke et al., 2014) to assess illusory self-touch ("I felt as if I was touching my back by myself"), somatic passivity ("I felt as if someone else was touching my back") and the feeling of a presence ("I felt as if someone was standing close to me"). The questionnaire contained five additional items related to the perception of vocal stimuli, which are reported in the supplementary material.

3.2.3.5   Respiration and heartbeat

Respiration and heartbeat signals were collected using a respiration belt and bipolar ECG electrodes (Biopac MP36R system), respectively, at a sampling rate of 2000 Hz. Respiration belt was placed about 5 cm below participants' armpits, whereas the bipolar electrodes were placed on the 2 clavicles and lower left rib. A trigger was sent at the onset of the second word in each word pair (Figure 1), in order to determine, for each trial, in which part of the heartbeat and respiration cycle the auditory stimulus occurred.

Cycle of the continuous respiration signal was divided into inspiration and expiration periods. We first obtained signal phase values by applying Hilbert transform to the bandpass-filtered signal between 0.2 and 0.8 Hz. Phase values belonging to the interval $(-\pi, 0)$ were classified as expiration, whereas those in the interval $(0, \pi)$ as inspiration. Systole and diastole of a heartbeat signal were defined as parts of a heartbeat cycle with previously defined onsets and durations relative to R peak (Kunzendorf et al., 2019). Preprocessing of both physiological signals was conducted using the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011).

3.2.3.6   Statistical analysis

The effect of respiration on performance in both auditory tasks was analyzed with mixed-effects binomial regressions with Response as dependent variable and Respiration (inspiration, expiration), Condition (synchronous, asynchronous) and Stimulus (levels: 1-6), together with a three-way interaction, as fixed effects. The Response-variable indicates whether participants perceived a stimulus as sounding more like their own voice (self-other task) or as louder (loudness task) compared to the reference stimulus. Random effects included a by-subject random intercept. By-subject random slopes for the main effects were added following model selection based on maximum likelihood. Trials with reaction times greater or smaller than two interquartile ranges from the median for each subject were considered as outliers and excluded.

The same mixed-effects binomial regression was applied to investigate the effect of heartbeat on auditory task performance, except that instead of the Respiration variable, the model contained Heartbeat variable with two levels: systole and diastole. A linear mixed-effects regression with Reaction Times as a dependent variable and the same fixed and random effects was also performed for both auditory tasks and both physiological signals.

The effect of Condition (synchronous, asynchronous) on ratings in questionnaire items was assessed by one-tailed t-tests, as the direction of the effect is known from previous work (Blanke et al., 2014; Salomon et al., 2020). For the questionnaire items that significantly differed between the two sensorimotor conditions (synchronous, asynchronous) on the group level, we further created a variable (e.g. Passivity variable), indicating whether individual participants experienced the illusion assessed by the corresponding question. Thus, participants were divided in two groups – those with a positive asynchronous-synchronous rating difference (e.g. Passivity+) and those with a negative or zero difference (e.g. Passivity-). Such variables were added as an additional fixed factor in the mixed-effects logistic regression assessing auditory task performance, to investigate potential effects of the subjective experience evoked by sensorimotor stimulation. Therefore, these regressions contained dependent variable Response and fixed effects of Respiration (inspiration, expiration), Stimulus (levels 1-6) and Illusion (+, -).

Finally, to control for the effects of sensorimotor stimulation and the auditory tasks on the measured physiological signals, we computed each participant's respiration and heartbeat rate and rate variability and for all four parameters performed a two-way ANOVA with Condition (synchronous, asynchronous) and Task (self-other, loudness) as fixed effects with an interaction term and by-subject random effects. Heart rate variability was represented as root mean square of successive RR interval differences (RMSSD) (Shaffer & Ginsberg, 2017), and respiration rate variability as a coefficient of variation (CV) (Noto, Zhou, Schuele, Templer, & Zelano, 2018).

To investigate potential confounds due to gender (there were more female participants), we repeated the main mixed-effects regressions with an additional main effect of Gender. However, there was no effect of Gender nor an interaction with other main effects (supplementary material) and no differences to the main results. Respiration parameters were computed using BreathMetrics (Noto et al., 2018) and heartbeat using BioSig (Schölgl, Vidaurre, & Sander, 2011) toolbox. Statistical tests were performed with R, using notably the lme4 (Bates, Mächler, Bolker, & Walker, 2015), lmerTest (Kuznetsova, Brockhoff, & Christensen, 2018), and afex (Singmann, Bolker, Westfall, & Aust, 2019). The results were illustrated in R using sjplot (Lüdecke, 2018) and ggplot2 (Wickham, 2016) packages.

3.2.4   Results

3.2.4.1   Respiratory phase

3.2.4.1.1   Self-other discrimination

A mixed-effects binomial regression assessing the dependence of Response to the effects of Respiration, Condition and Stimulus in the self-other discrimination task revealed a main effect of Respiration (estimate=-1.02, Z=-3.49, p<0.001), indicating a lower rate of 'self' response during inspiration, compared to expiration. We further observed a main effect of Stimulus (estimate=0.53, Z=4.84, p<0.001), showing that 'self' responses were more frequent when voice-morphs contained more self-features. The effect of Respiration significantly interacted with the effect of Stimulus (estimate=0.24, Z=3.17, p=0.002), revealing a steeper slope for the curve fitted for the inspiration phase. This indicates that participants were better in discriminating their own voice from another person's voice during inspiration compared to expiration (Figure 28). By fitting four-parameter psychometric curves, we additionally showed that points of subjective equivalence (PSE) and right asymptotes did not differ between the two curves, indicating that respiration
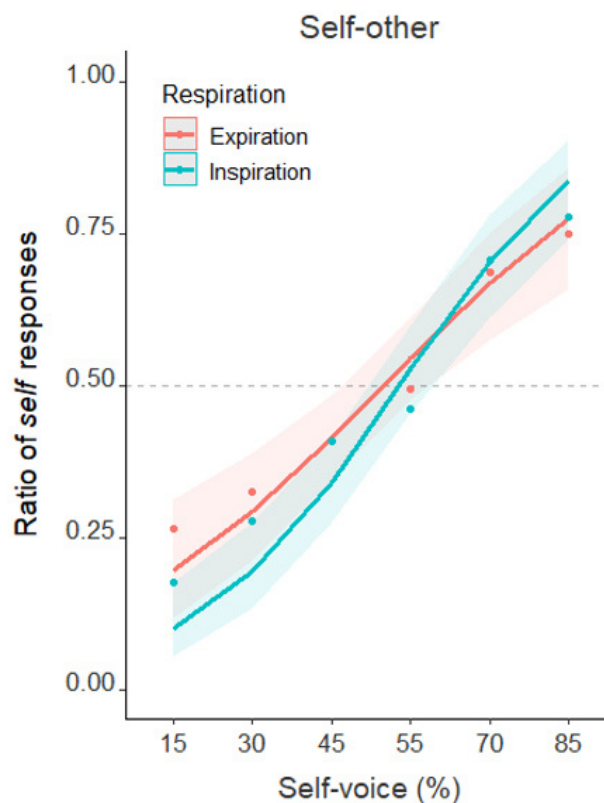
Figure 28. Breathing affects self-other voice discrimination.
Psychometric curves fitted for two respiration phases (expiration, inspiration) during the self-other task. Six stimulus levels on the abscissa represent six self-voice ratios and the ordinate indicates the rate at which the corresponding stimulus level was perceived as more resembling the 'self' than the baseline (50% self-voice). The dots represent grand average response. The shaded areas around each curve represent the 95% confidence intervals. A steeper curve fitting the perception during inspiration indicates that participants were better at discriminating between their own and someone else's voice during inspiration compared to expiration periods. This was especially prominent for other-dominant voice morphs.

did not introduce a bias in self-other voice discrimination and did not significantly improve recognition of self-dominant stimuli (supplementary material).

The mixed-effects binomial regression also revealed a two-way interaction between the effects of Respiration and Condition (estimate=0.89, Z=2.17, p=0.03) and a three-way interaction between the effects of Respiration, Condition and Stimulus (estimate=-0.21, Z=-1.97, p=0.048). To further investigate the nature of these interactions, we performed separate mixed-effects logistic regression for the two levels of Condition (synchronous and asynchronous). For the dataset containing the asynchronous experimental blocks, the effect of Respiration was significant (estimate=-0.98, Z=-3.30, p<0.001) and it significantly interacted with the effect of Stimulus (estimate=0.22, Z=2.92, p=0.004), again indicating a lower intercept and a steeper slope for the inspiration phase (Figure 3, left). On the contrary, such an effect of Respiration did not occur during synchronous sensorimotor stimulation (estimate=-0.08, Z=-0.27, p=0.79) nor did it interact with the effect of Stimulus (estimate=0.02, Z=0.22, p=0.83) (Figure 29, right). This analysis shows that the observed effects of the respiration on the self-other discrimination (Figure 28) were only found during asynchronous sensorimotor stimulation (Figure 29, left) – i.e. participants were better in discriminating their own from another person's voice during the asynchronous sensorimotor stimulation, but not during the synchronous stimulation. Equivalent models relying on the Bayesian framework revealed vidence in favor of the null hypothesis according to which breathing did not affect self-other discrimination in the synchronous condition (Bayes factor = 0.17, see supplementary material).
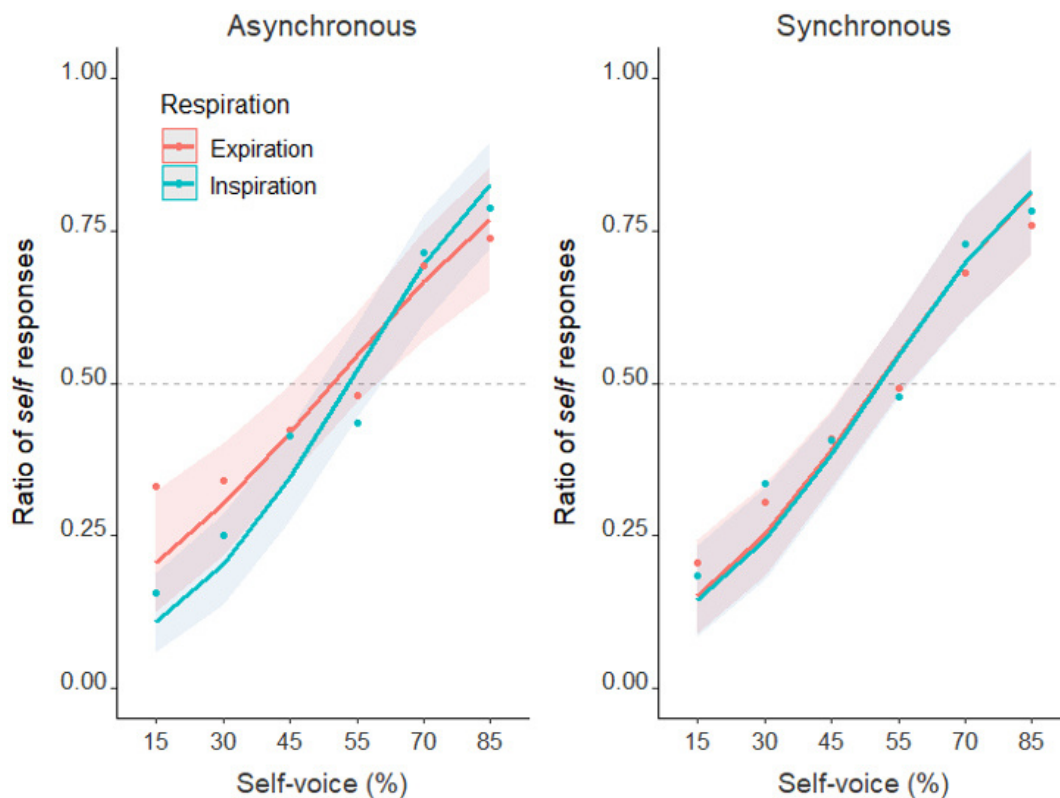
Figure 29. Breathing, self-other voice discrimination and sensorimotor stimulation.
Psychometric curves fitted for respiration phases for the two sensorimotor conditions of the self-other task. Inspiration was advantageous for self-other voice discrimination only during the asynchronous stimulation.

3.2.4.1.2   Bodily self-consciousness

As we reported previously (Orepic et al., 2020), the analysis of subjective ratings revealed that participants experienced stronger somatic passivity in the asynchronous (mean rating = 3.21, 95% CI = [2.34, 4.09]) versus synchronous (2.43, [1.57, 3.28]) condition (t(27)=2.05, p=0.025, Cohen's d=0.35). Participants rated illusory self-touch significantly stronger in the synchronous (2.39, [1.58, 3.21]) versus asynchronous (1.39, [0.61, 2.18]) condition (t(27)=2.58, p=0.008, d=0.49). There were no significant differences between conditions in other questionnaire items (all p>0.05).

To further assess the relationship between the observed differences in subjective experience and in auditory perception, we ran the same mixed-effects logistic regression on self-other task performance with an additional binary independent variable reflecting whether participants experienced somatic passivity and self-touch (see methods). Thus, participants were divided in two groups – those with a positive asynchronous-synchronous rating difference (Passivity+, N = 16 / Self-touch+, N = 14) and those with a negative or zero difference (Passivity-, N = 12 / Self-touch-, N = 14). We observed a significant interaction between Passivity and Respiration (estimate=-1.08, Z=-2.76, p=0.006) and a three-way interaction between Respiration, Stimulus and Passivity borderlined with significance (estimate=0.17, Z=1.73, p=0.084; other details of the model in the supplementary material). Investigation of these interactions revealed a steeper curve for the inspiration phase only in the Passivity+ group (Figure 30, left) (Respiration: estimate=-1.18, Z=-4.15, p<0.001; Stimulus: estimate=0.54, Z=3.94, p<0.001; Respiration-Stimulus interaction: estimate=0.24, Z=3.32, p<0.001). No such effects were found in the Passivity- group (Figure 30, right) (Respiration: estimate=0.13, Z=0.46, p=0.646; Stimulus: estimate=0.47, Z=9.65, p<0.001; Respiration-Stimulus interaction: estimate=0, Z=0.02, p=0.984). There were no significant interactions between Self-touch and Respiration (supplementary material).
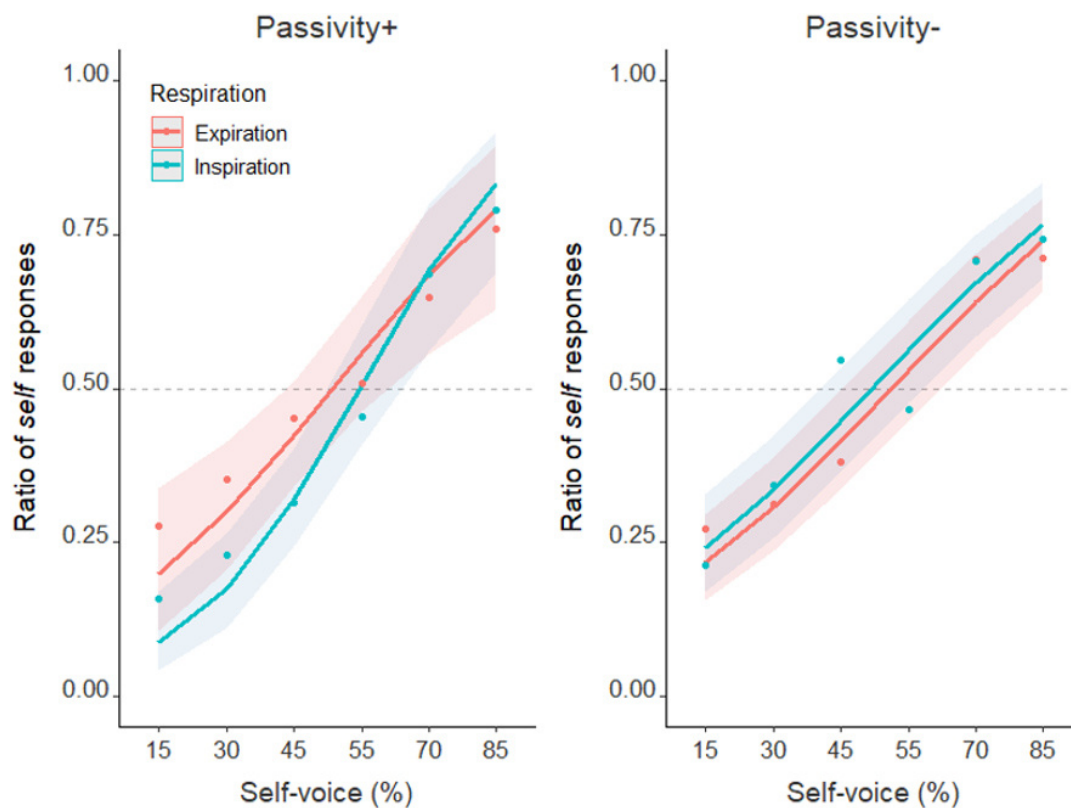
Figure 30. Breathing, self-other voice discrimination and somatic passivity.
Only participants experiencing somatic passivity (Passivity+, left) were better in discriminating between self and other voices during inspiration compared to expiration. Respiration did not affect voice perception in the Passivity- group (right).

3.2.4.1.3   Loudness perception

Mixed-effects binomial regression on loudness perception with Response as a dependent variable and Respiration, Condition and Stimulus as fixed effects revealed only a main effect of Stimulus (estimate=0.56, Z=11.73, p<0.001). Respiration did not affect loudness judgement (estimate=0.07, Z=0.28, p=0.780), nor did it interact with any of the other fixed effects (Condition: estimate=-0.25, Z=-0.68, p=0.497; Stimulus: estimate=0, Z=0.03, p=0.977). There was no main effect of Condition (estimate=-0.19, Z=-0.74, p=0.460), no two-way interaction between Condition and Stimulus (estimate=0.07, Z=1.01, p=0.310), nor a three-way interaction between Respiration, Condition and Stimulus (estimate=0.02, Z=0.22, p=0.825). These results suggest that loudness judgement does not depend on the respiration phase (Figure 31). Equivalent Bayesian models revealed evidence in favor of the null hypothesis according to which breathing did not affect loudness judgments (BF = 0.15, see supplementary material).
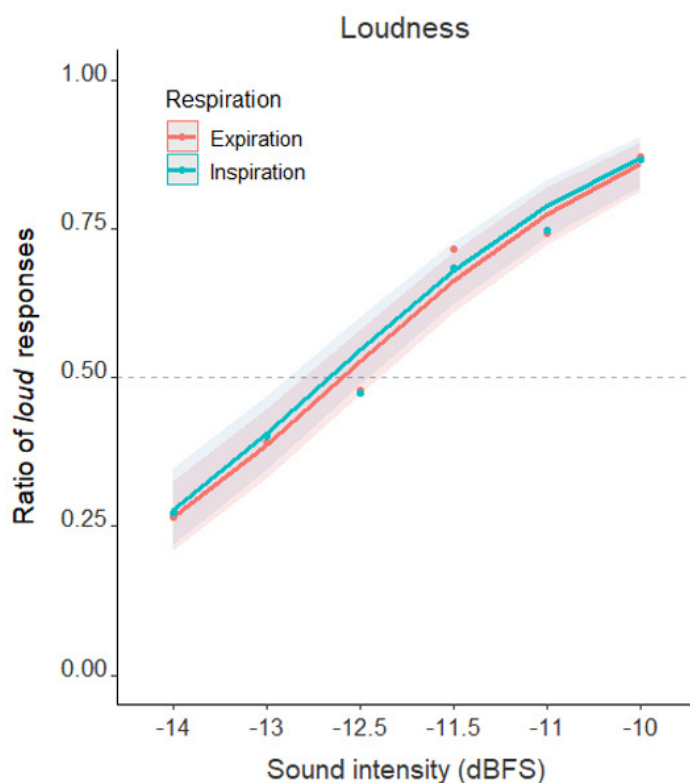
Figure 31. Breathing and loudness perception.
Psychometric curves fitted for two respiration phases (expiration, inspiration) during the loudness task. Six stimulus levels on the abscissa represent six sound intensity levels and the ordinate indicates the rate at which the corresponding stimulus level was perceived louder than the baseline (12 dBFS). The dots represent grand averages and the shaded areas around each curve represent the 95% confidence intervals. Respiration did not affect loudness judgement.

3.2.4.2   Cardiac phase

Heartbeat phase did not affect task performance in either of the auditory tasks – there were no main effects of Heartbeat either on self-other (estimate=0.26, Z=0.79, p=0.428), nor on loudness judgement (estimate=-0.27, Z=-0.92, p=0.359). Both models indicated a main effect of Stimulus (self-other: estimate=0.56, Z=10.08, p<0.001; loudness: estimate=0.53, Z=10.12, p<0.001) and no effect of Condition (self-other: estimate=0.21, Z=0.69, p=0.488; loudness: estimate=-0.49, Z=-1.73, p=0.083). There were no significant inter-actions between the fixed effects in either task (for a detailed report see supplementary material). Equivalent Bayesian models revealed evidence in favor of the null hypothesis according to which heartbeat did not affect self-other discrimination (BF = 0.17) nor loudness judgments (BF = 0.26, supplementary material).

3.2.4.3   Reaction times

There were no significant effects of Respiration on reaction times in either auditory task (loudness: estimate=0, t(105.1)=-0.2, p=0.839; self-other: estimate=0.01, t(3194)=0.79, p=0.432). Similarly, we observed no main effect of Hearbeat in either auditory task (loudness: estimate=0, t(2491)=-0.03, p=0.976; self-other: estimate=0.02, t(2423)=0.92, p=0.358). Finally, neither of the two effects (Respiration, Heartbeat) showed a significant interaction with other fixed effects (supplementary material).

3.2.4.4   Physiological analysis

Neither robotic stimulation nor the auditory tasks modulated parameters of respiratory or cardiac physiological signals. ANOVA did not indicate significant differences between respiration rate during asynchronous (mean = 0.35 Hz, 95% CI [0.33, 0.37] Hz) and synchronous (0.35, [0.33, 0.37] Hz) conditions (F(1, 26)=0.06, p=0.813) and also did not differ during loudness (0.35, [0.33, 0.37] Hz) and self-other (0.35, [0.33, 0.37] Hz) tasks (F(1, 26)=0.12, p=0.733). Similarly, respiration rate variability did not differ between these conditions (asynchronous (0.37, [0.31, 0.43] Hz) vs. synchronous (0.36, [0.30, 0.42] Hz) conditions (F(1, 26)=0.21, p=0.649); loudness (0.36, [0.30, 0.42] Hz) vs. self-other (0.37, [0.31, 0.44] Hz) tasks (F(1, 26)=0, p=0.964)). This was also found for heart rate (asynchronous (80.95, [77.36, 84.55] bpm) vs. synchronous (81.16, [77.81, 84.51] bpm) conditions (F(1, 26)=0, p=0.945); loudness

(80.59, [77.20, 83.99] bpm) vs. self-other (81.53, [77.98, 85.07] bpm) tasks (F(1, 26)=1.81, p=0.190)) and heart rate variability (asynchronous (36.55, [30.53, 42.57] bpm) vs. synchronous (35.09, [28.44, 41.73] bpm) (F(1, 26)=0.12, p=0.731); loudness (36.18, [29.25, 43.12] bpm) vs. self-other: 35.43, [29.75, 41.11] bpm) (F(1, 26)=0.13, p=0.717)). Finally, there were no significant interactions between the effects of Condition and task for any of the parameters (breathing rate: F(1, 26)=0.62, p=0.437; breathing rate variability: F(1, 26)=0.02, p=0.894; heart rate: F(1, 26)=0, p=0.999; heart rate variability: F(1, 26)=0.22, p=0.645).

3.2.5   Discussion

We report that participants were better in discriminating their own from someone else's voice during inspiration compared to expiration. Moreover, this inspiratory advantage for self-voice processing was stronger during the asynchronous sensorimotor stimulation and was thus more pronounced in the condition inducing illusory misattribution of self-generated sensations to someone else. Breathing did not affect auditory perception when participants made loudness judgements of the same self-related vocal stimuli and cardiac phase did not modulate the performance in either auditory task.

An advantage of inspiration over expiration in self-other discrimination has not been reported before and extends previous respiratory phase dependency data to self-related cognition. Previously, other cognitive processes have been shown to be improved during inspiration: memory retrieval (Zelano et al., 2016), spatial perception (Perl et al., 2019), visual pattern recognition (Nakamura et al., 2018), and emotion discrimination (Zelano et al., 2016). It has been argued that the inspiratory phase of the respiratory cycle drives neural synchronization of cortical and sub-cortical regions, thereby affecting the corresponding task-related neural activations (Heck et al., 2019; Zelano et al., 2016). Based on differences in resting-state functional connectivity between inspiration and expiration phases, it has also been proposed that inspiration-driven patterns of neural activity may improve the processing of incoming stimuli (Perl et al., 2019). Our results demonstrate a solid inspiratory-phase advantage based on psychophysics and thereby associate breathing to self-related processing. We observed that inspiration led to systematic sensitivity changes, but no perceptual bias, in self-other voice discrimination (i.e. steeper psychometric curve, Figure 2) and that it was driven by differences in other-dominant voice morphs (left asymptote; Figure 2). It is further worth noticing that the breathing effect was observed only for self-other voice discrimination while basic breathing parameters such as breathing rate and variability were equal in both auditory tasks. These data demonstrate that inspiration leads to improved and unbiased self-other voice discrimination and that it does not improve loudness judgement of the same vocal stimuli, showing that inspiration only impacted aspects of auditory perception related to discriminating one's own voice from another person's voice.

Another important finding was that the improvement in self-other discrimination during inspiration was driven by changes related to online sensorimotor stimulation, and, in particular, to asynchronous sensorimotor stimulation, with basic breathing parameters (rate and rate variability) being unaffected by the stimulation itself. Thus, the observed increase in sensitivity for self-other discrimination during the inspiration phase (as indicated by a steeper psychometric curve) was specific to performance during asynchronous sensorimotor stimulation. These differences between inspiration and expiration were absent in the synchronous condition, suggesting that the stronger sensorimotor conflict impacts the effect of the breathing cycle on self-other discrimination. Namely, asynchronous stimulation contains a stronger sensorimotor conflict, as there is an additional temporal conflict between poking movements in the front and tactile sensations on participants' back, in addition to the spatial conflict, which is also present during synchronous stimulation. Such robotically-applied sensorimotor conflicts during asynchronous stimulation have been related to changes in BSC, especially to the feeling of a presence and to somatic passivity (Blanke et al., 2014; Salomon et al., 2020). The present data link the inspiration-driven advantage for self-other voice discrimination not only to asynchronous sensorimotor simulation, but also to the altered BSC state associated with it. Thus, additional analysis revealed that the inspiration-driven effect was observed only in the participants reporting somatic passivity – i.e. in those participants susceptible to misattributing self-generated tactile sensations to someone else. This links the present breathing effect in self-other discrimination not only to the online respiratory cycle and sensorimotor state of the participant, but also to conscious self-representation (i.e. somatic passivity). Interestingly, the advantage of inspiration in self-other voice discrimination was again most pronounced for other-dominant voice morphs, indicating that in an other-oriented BSC state (i.e. otherness associated during asynchronous stimulation with somatic passivity) participants were more sensitive to recognizing self-other morphs as another person. We argue that respiration and conflicting sensorimotor stimulation combine to induce systematic changes in conscious self-other experience and self-other discrimination that depend on the respiration cycle. Our findings show an improvement in self-other voice discrimination, especially while hearing the voices of another person, in individuals experiencing a loss of self and enhanced otherness, suggesting that breathing facilitates self-other discrimination only during a BSC state characterized by otherness, a shift from self towards other.

To summarize, we demonstrate (1) a relationship between breathing and self-other voice discrimination, which is (2) dependent on sensorimotor integration and (3) related to feelings of otherness in the form of somatic passivity. From the two tested interoceptive functions, only respiration, but not cardiac, phase affected self-voice perception. Breathing is fundamentally related to speech and voice production (thus to the sound of our own voice) (Von Euler, 2011) and with voluntary action (H. Park et al., 2020). We argue that the present findings about the coupling between breathing and self-other voice discrimination may reflect that voice perception and the voluntary action of speaking are coupled with the basic physiological function of breathing, which are absent (or less pronounced) for cardiac physiology. We also did not observe cardiac-dependent differences in reaction times as it has been reported for self-face perception (Ambrosini et al., 2019), arguing that different physiological signals (e.g. respiration and heartbeat) affect self-related processes differently, depending on their intrinsic cyclic differences, their specific functional associations, and likely the investigated sensory modality. Our data shed new light on the interactions between interoception, BSC and self-voice perception and as such extend previous findings on breathing-dependent cognition to self-related processing.

Author contribution

Study concept and design: PO, GR, NF, OB. Acquisition of data: PO. Analysis and interpretation of data: PO, HP, GR, NF, OB. Drafting of the manuscript: PO, HP, NF, OB. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: PO, NF. Obtained funding: OB. Administrative, technical, or material support: All authors. Study supervision: NF, OB.

Open Practices Statement

Neither of the tasks reported in this article was formally preregistered. Anonymized data and analysis scripts are available on the open science framework (https://osf.io/kvqrn/).

3.2.6   References

Adler, D., Herbelin, B., Similowski, T., & Blanke, O. (2014). Breathing and sense of self: Visuo-respiratory conflicts alter body self-consciousness. Respiratory Physiology & Neurobiology, 203, 68–74. https://doi.org/10.1016/j.resp.2014.08.003

Ambrosini, A. E., Finotti, G., Ruben, T., Tsakiris, M., & Ferri, F. (2019). Seeing myself through my heart: cortical processing of a single heartbeat speeds up self-face recognition. Biological Psychology, 44(200), 1–30. https://doi.org/10.1093/annonc/mdy039/4835470

Aspell, J. E., Heydrich, L., Marillier, G., Lavanchy, T., Herbelin, B., & Blanke, O. (2013). Turning Body and Self Inside Out: Visualized Heartbeats Alter Bodily Self-Consciousness and Tactile Perception. Psychological Science, 24(12), 2445–2453. https://doi.org/10.1177/0956797613498395

Azevedo, R. T., Garfinkel, S. N., Critchley, H. D., & Tsakiris, M. (2017). Cardiac afferent activity modulates the expression of racial stereotypes. Nature Communications, 8, 1–9. https://doi.org/10.1038/ncomms13854

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software. https://doi.org/10.18637/jss.v067.i01

Birren, J. E., Cardon, P. V., & Phillips, S. L. (1963). Reaction time as a function of the cardiac cycle in young adults. Science, 140(3563), 195–196. https://doi.org/10.1126/science.140.3563.195-a

Blanke, O., Pozeg, P., Hara, M., Heydrich, L., Serino, A., Yamamoto, A., … Rognini, G. (2014). Neurological and robot-controlled induction of an apparition. Current Biology, 24(22), 2681–2686. https://doi.org/10.1016/j.cub.2014.09.049

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness. Neuron, 88(1), 145–166. https://doi.org/10.1016/j.neuron.2015.09.029

Faivre, N., Vuillaume, L., Bernasconi, F., Salomon, R., Blanke, O., & Cleeremans, A. (2020). Sensorimotor conflicts alter metacognitive and action monitoring. Cortex, 124(22), 2681–2686. https://doi.org/10.1016/j.cortex.2019.12.001

Garfinkel, S. N., Minati, L., Gray, M. A., Seth, A. K., Dolan, R. J., & Critchley, H. D. (2014). Fear from the Heart: Sensitivity to Fear Stimuli Depends on Individual Heartbeats. Journal of Neuroscience, 34(19), 6573–6582. https://doi.org/10.1523/JNEUROSCI.3507-13.2014

Hara, M., Rognini, G., Evans, N., Blanke, O., Yamamoto, A., Bleuler, H., & Higuchi, T. (2011). A novel approach to the manipulation of body-parts ownership using a bilateral master-slave system. In IEEE International Conference on Intelligent Robots and Systems (pp. 4664–4669). https://doi.org/10.1109/IROS.2011.6048519

Heck, D. H., Kozma, R., & Kay, L. M. (2019). The rhythm of memory: how breathing shapes memory function. Journal of Neurophysiology, 122(2), 563–571. https://doi.org/10.1152/jn.00200.2019

Kawahara, H., Morise, M., Banno, H., & Skuk, V. G. (2013). Temporally variable multi-aspect N-way morphing based on interference-free speech representations. In 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2013 (pp. 1–10). IEEE. https://doi.org/10.1109/APSIPA.2013.6694355

Kleiner, M., Brainard, D. H., Pelli, D. G., Ingling, A., Murray, R., & Broussard, A Ingling, R Murray, C. (2007). What's new in Psychtoolbox-3? Perception, 36(14), 1–16. https://doi.org/10.1068/v070821

Kunzendorf, S., Klotzsche, F., Akbal, M., Villringer, A., Ohl, S., & Gaebler, M. (2019). Active information sampling varies across the cardiac cycle. Psychophysiology, 1–16. https://doi.org/10.1111/psyp.13322

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lüdecke, D. (2018). sjPlot: Data Visualization for Statistics in Social Science. R package version 2.6.2. https://doi.org/10.5281/zenodo.1308157

Motyka, P., Grund, M., Forschack, N., Al, E., Villringer, A., & Gaebler, M. (2019). Interactions between cardiac activity and conscious somatosensory perception. Psychophysiology, 56(10), e13424. https://doi.org/10.1111/psyp.13424

Nakamura, N. H., Fukunaga, M., & Oku, Y. (2018). Respiratory modulation of cognitive performance during the retrieval process. PLoS ONE, 13(9), e0204021.

Noto, T., Zhou, G., Schuele, S., Templer, J., & Zelano, C. (2018). Automated analysis of breathing waveforms using BreathMetrics: A respiratory signal processing toolbox. Chemical Senses, 43(8), 583–597. https://doi.org/10.1093/chemse/bjy045

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience. https://doi.org/10.1155/2011/156869

Orepic, P., Rognini, G., Kannape, O. A., Faivre, N., & Blanke, O. (2020). Sensorimotor conflicts induce somatic passivity and louden quiet voices in healthy listeners. BioRxiv. https://doi.org/https://doi.org/10.1101/2020.03.26.005843

Park, H., Barnoud, C., Trang, H., Kannape, O. A., Schaller, K., & Blanke, O. (2020). Breathing is coupled with voluntary action and the cortical readiness potential. Nature Communications, 11(1), 1–8. https://doi.org/10.1038/s41467-019-13967-9

Park, H. D., & Blanke, O. (2019). Coupling Inner and Outer Body for Self-Consciousness. Trends in Cognitive Sciences, 23(5), 377–388. https://doi.org/10.1016/j.tics.2019.02.002

Perl, O., Ravia, A., Rubinson, M., Eisen, A., Soroka, T., Mor, N., … Sobel, N. (2019). Human non-olfactory cognition phase-locked with inhalation. Nature Human Behaviour, 3(5), 501–512. https://doi.org/10.1038/s41562-019-0556-z

Pfeifer, G., Garfinkel, S. N., Gould van Praag, C. D., Sahota, K., Betka, S., & Critchley, H. D. (2017). Feedback from the heart: Emotional learning and memory is controlled by cardiac cycle, interoceptive accuracy and personality. Biological Psychology, 126, 19–29. https://doi.org/10.1016/j.biopsycho.2017.04.001

Salomon, R., Progin, P., Griffa, A., Rognini, G., Do, K. Q., Conus, P., … Blanke, O. (2020). Sensorimotor Induction of Auditory Misattribution in Early Psychosis. Schizophrenia Bulletin. https://doi.org/10.1093/schbul/sbz136

Sandman, C. A., McCanne, T. R., Kaiser, D. N., & Diamond, B. (1977). Heart rate and cardiac phase influences on visual perception. Journal of Comparative and Physiological Psychology, 91(1), 189. https://doi.org/10.1037/h0077302

Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: Discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. Cognition, 94(3), 241–255. https://doi.org/10.1016/j.cognition.2004.04.003

Schölgl, A., Vidaurre, C., & Sander, T. H. (2011). BioSig: The free and open source software library for biomedical signal processing. Computational Intelligence and Neuroscience. https://doi.org/10.1155/2011/935364

Seth, A. K., & Tsakiris, M. (2018). Being a Beast Machine: The Somatic Basis of Selfhood. Trends in Cognitive Sciences, 22(11), 969–981. https://doi.org/10.1016/j.tics.2018.08.008

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. Frontiers in Public Health, 5, 258. https://doi.org/10.3389/fpubh.2017.00258

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2019). afex: Analysis of Factorial Experiments. R package version 0.23-0. https://cran.r-project.org/package=afex. Retrieved from https://cran.r-project.org/package=afex

Suzuki, K., Garfinkel, S. N., Critchley, H. D., & Seth, A. K. (2013). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. Neuropsychologia, 51(13), 2909–2917. https://doi.org/10.1016/j.neuropsychologia.2013.08.014

Tsakiris, M., Haggard, P., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. Cognition, 96(3), 215–231. https://doi.org/10.1016/j.cognition.2004.08.002

Von Euler, C. (2011). Brain Stem Mechanisms for Generation and Control of Breathing Pattern. In Comprehensive Physiology (pp. 1–67). https://doi.org/10.1002/cphy.cp030201

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wilkinson, M., McIntyre, D., & Edwards, L. (2013). Electrocutaneous pain thresholds are higher during systole than diastole. Biological Psychology, 94(1), 71–73. https://doi.org/10.1016/j.biopsycho.2013.05.002

Zelano, C., Jiang, H., Zhou, G., Arora, N., Schuele, S., Rosenow, J., & Gottfried, J. A. (2016). Nasal Respiration Entrains Human Limbic Oscillations and Modulates Cognitive Function. The Journal of Neuroscience, 36(49), 12448–12467. https://doi.org/10.1523/JNEUROSCI.2586-16.2016

3.2.7 Supplementary material

3.2.7.1 Additional analyses

To further specify the nature of perceptual alterations caused by the respiration phase, for each participant and each respiration phase (inspiration, expiration) we estimated four parameters of the corresponding psychometric curve – point of subjective equivalence (PSE), slope, left and right asymptote – using quickpsy (Linares & López-Moliner, 2016) library. A two-way ANOVA with Parameter Value as a dependent variable and fixed effects Respiration (inspiration, expiration) and Parameter (PSE, slope, left and right asymptote) showed a significant interaction between the two effects (F(1.45, 39.06)=4.03, p=0.04). A two-tailed paired t-test was thus performed for each Parameter to assess whether its Parameter Value differed between inspiration and expiration phases. Consistent with the mixed-effects binomial regression, the curve fitted for the inspiration phase had a steeper slope (t(27)=2.23, p=0.034, Cohen's d=0.61) and a lower left asymptote (t(27)=-2.49, p=0.019, d=0.46). Additionally, there were no differences between the values of PSE (t(27)=0.61, p=0.546, d=0.14) and right asymptote (t(27)=0, p=0.995, d=0). These effects show that inspiration phase increased sensitivity to self-other voice discrimination (slope effect) and that the difference in performance was especially prominent for the other-dominant voice morphs (left asymptote) (main Figure 2). Respiration did not introduce a bias in self-other voice perception (absence of the PSE effect) and did not significantly improve recognition of self-dominant stimuli (right asymptote).

3.2.7.2 Bayesian analyses of the null effects

To further validate nulls findings of the mixed-effects binomial regressions reported in the main text, we ran three main regressions in a Bayesian framework: (1) for the absence of the respiration effect in loudness task, (2) for the absence of the respiration effect in the synchronous condition of the self-other task, and (3) for the absence of the cardiac effect. All Bayesian models were created in Stan computational framework (http://mc-stan.org/) accessed with the brms package (Bürkner, 2017), based on four chains of 10000 iterations including 2000 warmup samples. We report the highest density probability for all estimates, which specifies the range covering the 95% most credible values of the posterior estimates.

For the null effect of Respiration in the loudness task we had a prior assumption of better performance during the inspiration phase – represented by a steeper slope – (i.e., prior on the interaction between the effects of Respiration and Stimulus with Gaussian distribution of mean = 0.24 and SD = 0.4, based on the same interaction observed in the self-other task). No interaction between Respiration and Stimulus was found (estimate = 0.02, highest posterior density interval = [-0.12 0.15], Bayes factor = 0.15). Similar effects were observed even with a Gaussian prior centered on zero (estimate = 0.01, CI = [-0.13 0.14], BF = 0.17). Using the same prior (mean = 0.24, SD = 0.4), we observed no significant interaction between Respiration and Stimulus for the synchronous condition of the self-other task (estimate = 0.03, CI = [-0.11 0.17], BF = 0.17). Bayes factors smaller than 0.3 support the null hypothesis, according to which breathing did not affect loudness judgement nor self-other discrimination during the synchronous condition.

Similarly, for the cardiac phase, we had a prior assumption of a better performance during systole – represented by a steeper slope – (i.e., prior on the interaction between the effects of Heartbeat and Stimulus with Gaussian distribution of mean = 0.24 and SD = 0.4, based on the same interaction observed for the Respiration effect in the self-other task). No interaction between Heartbeat and Stimulus was found for either task (Self-other: estimate = 0, CI = [-0.16 0.15], BF = 0.17; Loudness: estimate = 0.08, CI = [-0.08 0.23], BF = 0.26). Again, Bayes factors smaller than 0.3 support the null hypothesis of no effects of heartbeat phase on auditory tasks' performance.

3.2.7.3 Words

As reported in our previous work (Orepic, Rognini, Kannape, Faivre, & Blanke, 2020), participants were recorded saying 10 words in French (clou, fouet, hache, lame, lutte, os, rat, sang, scie, ver). The words were chosen from the list of 100 negatively-valenced words, as rated by 20 schizophrenic patients and 97 healthy participants (Jalenques, Enjolras, & Izaute, 2013). Negative words were purposefully chosen in our previous study (Orepic et al., 2020), in order to better approximate the phenomenology of auditory-verbal hallucinations (AVH), that are mostly negative in content (Woods, Jones, Alderson-Day, Callard, & Fernyhough, 2015).

3.2.7.4 Bodily self-consciousness

An overview of the questionnaire, also reported in our previous work (Orepic et al., 2020) is given in Table S1.

Table 3. Questionnaire.

Likert-scale questionnaire used to assess subjective experience evoked by sensorimotor stimulation.

Self-touch	**I felt as if I was touching my back by myself.**
Somatic passivity	I felt as if someone else was touching my back.
Control	I felt as if I had 3 bodies.
Presence hallucination	I felt as if someone was standing close to me.
Friend-voice	I felt as if I heard my friend's voice more often than my own voice
Friend-presence	I felt as if my friend was standing close to me.
Other-presence	I felt as if someone else than my friend was standing close to me
Other-voice	I felt as if I heard a voice that was neither my friend's nor mine.
Bodily sensations	While hearing the words, I felt changes in my body sensations (e.g. lighter, warmer, I felt tingling sensations etc.)

As indicated and reported in the main text, for the self-other task, we ran two mixed-effects binomial regression with a dependent variable Response and fixed effects of Respiration, Stimulus and Passivity/Self-touch. Besides a significant interaction between Respiration and Passivity (see main text), the model which had Passivity as an additional factor only showed a significant effect of Stimulus (estimate=0.47, Z=9.67, p<0.001). The main effects of Passivity (estimate=0.12, Z=0.38, p=0.704) and of Respiration (estimate=0.13, Z=0.46, p=0.645) were not significant. Stimulus did not interact significantly neither with the effect of Respiration (estimate=0, Z=0.02, p=0.985), nor with the effect of Passivity (estimate=0, Z=-0.11, p=0.911). The model with Self-touch as an additional factor also showed a significant effect of Stimulus (estimate=0.5, Z=10.84, p<0.001), whereas effects of Respiration (estimate=-0.19, Z=-0.69, p=0.492) and Self-touch were not significant (estimate=0.47, Z=1.45, p=0.148). This model identified no significant interactions: Stimulus and Self-touch (estimate=-0.07, Z=-1.11, p=0.266), Respiration and Self-touch (estimate=-0.5, Z=-1.3, p=0.194), Stimulus and Respiration (estimate=0.07, Z=0.94, p=0.347), Respiration, Stimulus and Self-touch (estimate=0.05, Z=0.51, p=0.608).

3.2.7.5 Control analysis of the Gender effect

To investigate potential confounds due to gender (there were more female participants), we conducted a logistic mixed-effects regression with an additional main effect of Gender. As described in the main text, the model contained Response as a dependent variable and besides the effects of Respiration (inspiration, expiration), Condition (asynchronous, synchronous) and Stimulus (levels: 1-6) – all related with an interaction term – the newly added effect of Gender (male, female) was related with the effect of Respiration with an interaction term. As model having a four-way interaction between all the fixed effects did not converge, and because the effect of Respiration was of the main interest in the study, we chose to model its interactions with all other effects, including Gender. The model showed neither a main effect of Gender (estimate=0, Z=-0.02, p=0.983) nor a significant interaction between Gender and Respiration (estimate=0.13, Z=0.78, p=0.433).

3.2.7.6 Cardiac phase

In Table S2 we report the non-significant interactions between the fixed effects in both tasks, omitted from the main text, and in Figure 32, we illustrate the lack of cardiac phase effects on the performance of both auditory tasks.

Table 4. Model details for Heartbeat.

An overview of the interactions containing fixed effect of Heartbeat in binomial mixed-effects regressions assessing performance in Self-other and Loudness tasks.

Task	Heartbeat *	estimate	Z value	p value
Self-other	Condition	-0.4	-0.88	0.377
	Stimulus	-0.03	-0.32	0.752
	Condition * Stimulus	0.04	0.36	0.717
Loudness	Condition	0.39	0.92	0.359
	Stimulus	0.06	0.79	0.429
	Condition * Stimulus	-0.06	-0.48	0.63
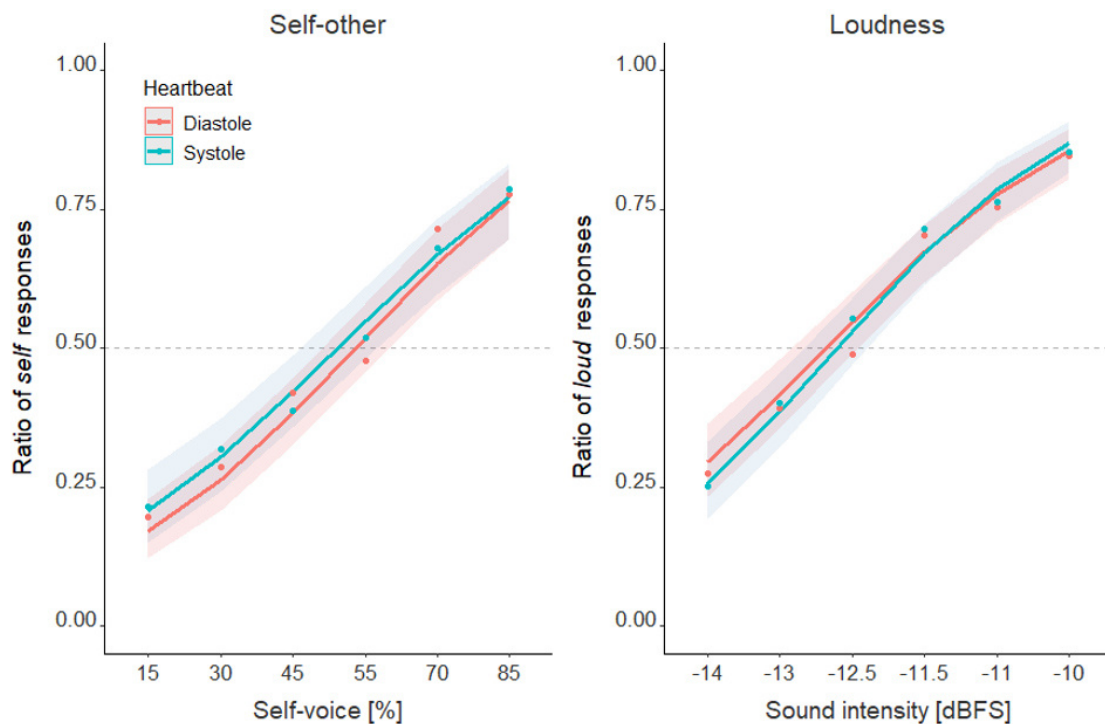
Figure 32. Heartbeat did not affect performance in auditory tasks.

Psychometric curves fitted for two heartbeat phases (systole, diastole) during self-other (left) and loudness (right) tasks. Six stimulus levels on the abscissa represent six self-voice ratios (left) and six sound intensity levels (right), whereas the ordinate indicates the rate at which the corresponding stimulus level was perceived as more resembling the 'self' (left) or as louder (right) than the baseline (50% self-voice, 12 dBFS). The shaded areas around each curve represent the 95% confidence intervals. Heartbeat did not affect performance in either auditory task.

3.2.7.7   Reaction times

As main effects of Respiration and Heartbeat were reported in the main text for both auditory tasks (self-other, loudness), here we report their interactions with other fixed effects, none of which was significant (Table 5).

Table 5. Model details for Reaction Times.

An overview of the interactions containing fixed effects of Respiration and Heartbeat in linear mixed-effects regressions assessing response times in Self-other and Loudness tasks.

Task	Interaction		estimate	df	t value	p value
Self-other	Respiration *	Condition	0.01	3203	0.56	0.577
		Stimulus	-1.24	3191	-0.95	0.342
		Condition * Stimulus	-0.39	3192	-0.21	0.832
	Heartbeat *	Condition	0	2428	-0.03	0.978
		Stimulus	-1.02	2416	-0.78	0.438
		Condition * Stimulus	1.15	2418	0.64	0.524
Loudness	Respiration *	Condition	0	3185	-0.26	0.797
		Stimulus	0.77	3152	0.83	0.408
		Condition * Stimulus	-0.44	3157	-0.33	0.739
	Heartbeat *	Condition	0	2486	0.13	0.897
		Stimulus	-0.11	2485	-0.12	0.907
		Condition * Stimulus	0.6	2484	0.42	0.673

3.2.7.8   Supplementary references

Bürkner, P. C. (2017). brms: An R package for Bayesian Multilevel Models using Stan. Journal of Statistical Software, 80(1).

Jalenques, I., Enjolras, J., & Izaute, M. (2013). Valence émotionnelle des mots. L'Encéphale, 39, 189–197.

Linares, D., & López-Moliner, J. (2016). quickpsy: An R Package to Fit Psychometric Functions for Multiple Groups. The R Journal, 8(1), 122–131. Retrieved from https://journal.r-project.org/archive/2016-1/linares-na.pdf

Orepic, P., Rognini, G., Kannape, O. A., Faivre, N., & Blanke, O. (2020). Sensorimotor conflicts induce somatic passivity and louden quiet voices in healthy listeners. BioRxiv. https://doi.org/https://doi.org/10.1101/2020.03.26.005843

Woods, A., Jones, N., Alderson-Day, B., Callard, F., & Fernyhough, C. (2015). Experiences of hearing voices: Analysis of a novel phenomenological survey. The Lancet Psychiatry, 2(4), 323–331. https://doi.org/10.1016/S2215-0366(15)00006-1

3.3 Study 5: Robotically-mediated sensorimotor stimulation induces identity-specific auditory-verbal hallucinations in healthy individuals

Pavo Orepic[1], Fosco Bernasconi[1], Melissa Faggella[1], Nathan Faivre[2], Olaf Blanke[1, 3]

1 Laboratory of Cognitive Neuroscience, Center for Neuroprosthetics and Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Switzerland

2 Laboratoire de Psychologie et Neurocognition (LPNC), CNRS UMR 5105, Université Grenoble Alpes, France

3 Faculty of Medicine, University of Geneva, Geneva, Switzerland

Corresponding author

Olaf Blanke

Bertarelli Chair in Cognitive Neuroprosthetics, Center for Neuroprosthetics & Brain Mind Institute, School of Life Sciences, Campus Biotech, Swiss Federal Institute of Technology (EPFL), 1012 Geneva, Switzerland

E-mail: olaf.blanke@epfl.ch

Tel: +41 (0)21 693 69 21

Fax: +41 (0)21 693 69 22

3.3.1 Abstract

Hearing voices without speakers present – denoted as auditory-verbal hallucinations (AVH) – is one of the most common as well as most distressing psychotic symptom, whose etiology is still largely unknown. As current treatments often prove ineffective, it is necessary to better understand its underlying principles. A prominent account portrays AVH as a deficit in auditory-verbal self-monitoring, characterized by a misattribution of inner speech towards an external agent talking to the individual. However, the empirical support relating self-monitoring impairments to AVH is solely correlational and causal evidence employing experimentally-induced self-monitoring impairments that would lead to AVH is still lacking. Recently, we introduced a robotic procedure able to induce selective deficits in bodily self-monitoring by applying sensorimotor conflicts of various degree between upper-limb movements and the corresponding tactile sensations on the back. In two independent studies (N1 = N2 = 24), we investigated whether two distinct robotic-sensorimotor stimulations, associated with increases in self- and other-agency sensations, can induce AVH in healthy individuals. AVH were quantified as a false alarm rate in a voice detection task employing self- and other vocal stimuli presented at individual hearing thresholds. In both studies, we observed and replicated an increase in self- and other-voice false alarms during the corresponding sensorimotor stimulation. False vocal percepts were further positively related to delusion proneness. By demonstrating an experimental procedure able to induce AVH in healthy individuals in a controlled laboratory environment, we shed new light on the phenomenology of AVH, relating them to bodily self-monitoring and delusional ideation.

Keywords

Auditory-verbal hallucinations, sensorimotor conflicts, self-monitoring, voice detection, self-voice, false alarms, delusion proneness

3.3.2 Introduction

Auditory-verbal hallucinations (AVH) – the sensation of hearing voices without speakers present – are one of the most common (Bauer et al., 2011) and most distressing (Harkavy-Friedman et al., 2003) symptoms in schizophrenia spectrum disorder, with contemporary treatments being effective only to a certain degree (Lehman et al., 2004). Moreover, AVH have been observed as well in non-treatment-seeking individuals (Powers, Kelley, & Corlett, 2017; Sommer et al., 2010) and characterized by a heavily heterogeneous phenomenological experience (e.g. varying with respect to voice numerosity, gender, frequency, emotional affect, etc.) (McCarthy-Jones et al., 2014; Woods, Jones, Alderson-Day, Callard, & Fernyhough, 2015). Thus, understanding the mechanisms leading to AVH is a critical next step towards the development of new treatments that are more soundly based upon systems neuroscience and brain pathophysiology.

Despite the frequency and the amount of studies revolving around AVH, the brain mechanisms underlying AVH still remain unknown. A prominent account suggests that AVH arise as a deficit in self-monitoring, whose main purpose is to facilitate distinguishing self- from other-generated stimuli. According to the self-monitoring framework (Miall & Wolpert, 1996; Shadmehr, Smith, & Krakauer, 2010; Wolpert, Ghahramani, & Jordan, 1995), self-other distinction is achieved by creating sensory predictions of own actions and by comparing them with the actual sensory feedback following those actions. When congruent with the prediction, sensory feedback is attenuated, and the action is attributed to the self, whereas if incongruent, there is no attenuation and the action is attributed to another external agent. Deficits in self-monitoring have been observed in schizophrenia (Blakemore, Smith, Steel, Johnstone, & Frith, 2000; Shergill, Samson, Bays, Frith, & Wolpert, 2005; Shergill et al., 2014), and therefore related to various psychotic symptoms characterized by a misattribution of self-generated actions towards external agents, such as passivity experiences and delusions of control (Feinberg, 1978; Frith, 1987; Frith, Blakemore, & Wolpert, 2000). Accordingly, placed under the umbrella of self-monitoring deficits, AVH have been portrayed as a self-to-other misattribution of inner speech (Ford, Roach, Faustman, & Mathalon, 2007; Frith, 1992; Frith & Done, 1989; Moseley, Fernyhough, & Ellison, 2013; Shergill et al., 2003), resulting from erroneous feedforward mechanisms associated to speaking.

The self-monitoring account for AVH was originally built upon the reports of EMG activity and motor cortex activation occurring both during AVH (Gould, 1948; Green & Kinsbourne, 1990; Green & Preston, 1981; McGuigan, 1966) and inner speech (Jacobson, 1931; Livesay, Liebke, Samaras, & Stanley, 1996; McGuigan & Dollins, 1989; Wildgruber, Ackermann, Klose, Kardatzki, & Grodd, 1996), suggesting that they may serve as motor actions. Empirical support to this assumption mainly comes from studies (reviewed by (Whitford, 2019)) in which patients with schizophrenia exhibited a reduced suppression of auditory evoked response while speaking compared to passively hearing their voice, as well as from reports of differences in functional connectivity (Ford, Math-

alon, Whitfield, Faustman, & Roth, 2002; Hoffman, Fernandez, Pittman, & Hampson, 2011), both of which were proposed to facilitate erroneous feedforward signaling. Crucially, the putative relationship between self-monitoring impairments and AVH is solely inferred upon correlational evidence – relating behavioral effects to symptom severity (which is often not replicated (Whitford, 2019)). Thus, causal evidence employing experimentally-induced self-monitoring impairments that would lead to AVH is still lacking.

Recently, we introduced a robotic procedure that can induce mild hallucinations in healthy individuals by perturbing bodily self-monitoring mechanisms (Bernasconi et al., 2020; Blanke et al., 2014; Faivre et al., 2020; Orepic, Rognini, Kannape, Faivre, & Blanke, 2020; Salomon et al., 2020). Specifically, we designed a two-part robotic setup that creates sensorimotor conflicts between repeated upper-limb poking movements and the corresponding tactile sensations on the back (Hara et al., 2011). With such a setup, sensorimotor conflicts of various degree can be facilitated through two types of sensorimotor stimulation – synchronous and asynchronous. Synchronous sensorimotor stimulation, that consists of a spatial conflict between the poking movements (extended to the front of the body) and the corresponding touches (perceived on the back), has been associated with illusory self-touch sensations, characterized by a self-attribution of touches despite the spatial incongruence with the movement. By adding an additional temporal conflict (i.e. a delay) between the movement and the corresponding touches (asynchronous stimulation), we were able to induce sensations of somatic passivity – the impression that tactile sensations are being applied/generated by an external agent and the feeling of external agent although no body is present (presence hallucination). Both of this phenomena are experienced frequently and regularly by psychotic patients. Asynchronous stimulation (mediating spatiotemporal sensorimotor conflicts) was associated with an increase in other-agency sensations, whereas synchronous stimulation (spatial conflicts) led to illusory self-agency.

Building up on self-monitoring accounts for AVH, we conducted two studies with the purpose of extending robotically-induced impairments in bodily self-monitoring to auditory-verbal processing. First, (Salomon et al., 2020) showed that asynchronous robotic stimulation can induce impairments in auditory self-monitoring in patients with passivity experiences, characterized by a decreased accuracy in auditory-verbal self-other discrimination. Second, (Orepic, Rognini, Kannape, Faivre, & Blanke, 2020) extended the findings of (Salomon et al., 2020) and related robotically-mediated sensorimotor stimulation to impairments in voice perception in a healthy population. Moreover, these effects were further associated to autonomic processing, demonstrating a dependence of auditory-verbal self-monitoring on the respiration phase (Orepic, Park, Rognini, Faivre, & Blanke, 2020). These studies showed that sensorimotor stimulation can cause cross-modal effects on auditory perception, however, it remains unknown whether it can directly induce auditory hallucinations.

Inspired by the associations between hallucinations and increased false alarm rates in auditory detection tasks (Barkus et al., 2011; Moseley, Fernyhough, & Ellison, 2014; Powers et al., 2017), here we extended our hallucination-inducing robotic procedure to a voice detection paradigm and investigated whether a specific sensorimotor stimulation pattern could cause an increase in vocal false alarms. Specifically, while exposed to synchronous and asynchronous sensorimotor stimulations, participants were performing a voice detection task, during which they were hearing self- or other voices at their hearing threshold, embedded in noise. We expected to observe a higher false alarm rate during asynchronous stimulation, that has been previously associated with hallucinations in healthy individuals (Blanke et al., 2014). Additionally, as asynchronous stimulation has been associated to somatic passivity and presence hallucination, both of which introduce other-agency sensations, we predicted that the increase in vocal false alarms would be pronounced with voices other than self. Finally, we conducted correlation analyses between vocal false alarm rates delusion-proneness (Peters, Joseph, Day, & Qarety, 2004) that has been related to deficits in self-monitoring (Teufel, Kingdon, Ingram, Wolpert, & Fletcher, 2010), allowing us to further explore the nature of potential effects, by investigating the relationship between the degrees of self-monitoring impairments, experimentally-induced auditory hallucinations and delusional ideation.

3.3.3 Method

3.3.3.1 Participants

We conducted two studies with the same general procedure and experimental design. Study 2 was set to replicate the effects observed in Study 1. Both studies involved 24 right-handed participants chosen from the general population, fluent in French and naïve to the purpose of the study. In Study 1, 17 participants were female (mean age ± SD: 25.0 ± 4.2 years old), whereas in Study 2, 13 were female (26.6 ± 5.3 years old). Sample size in both studies was determined to match the number of all possible permutations of experimental conditions. No participants reported any history of psychiatric or neurological disorders as well as any hearing

deficits. Participants gave informed consent in accordance with the institutional guidelines (protocol 2015-00092, approved by the Comité Cantonal d'Ethique de la Recherche of Geneva), and received monetary compensation (CHF 20/h).

3.3.3.2 Stimuli

Participants' voices were recorded (Zoom H6 Handy recorder) while saying nine one-syllable words in French (translated to English: nail, whip, ax, blade, fight, bone, rat, blood, saw, worm). The words were chosen from the list of 100 negatively-valenced words, as rated by 20 schizophrenic patients and 97 healthy participants (Jalenques, Enjolras, & Izaute, 2013). Negative words were purposefully chosen in our previous study (Orepic et al., 2020), in order to better approximate the phenomenology of AVH, that are mostly negative in content (Woods et al., 2015). After the background noise was removed from the recordings, they were standardized for sound intensity (-12 dBFS) and duration (500 milliseconds) (Audacity software). The preprocessed recordings were used as self-voice stimuli in a voice detection task, which also contained other-voice stimuli – i.e. equivalent voice recordings of a gender-matched person unknown to the participant. Auditory stimuli were presented to participants through noise-cancelling headphones (Bose QC20). The experimental paradigm was created in MATLAB 2017b with Psychtoolbox library (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

3.3.3.3 Experimental procedure

Upon arrival, participants were screened for eligibility criteria, after which their voices were recorded. This was followed by two Sensorimotor blocks (synchronous and asynchronous), designed to assess illusory effects of sensorimotor stimulation. Sensorimotor blocks were followed by Staircase blocks (bottom-up and top-down), used to estimate individual hearing thresholds with a voice detection task. Finally, in four Task blocks (synchronous-self, synchronous-other, asynchronous-self, asynchronous-other) we assessed vocal false alarms by combining sensorimotor stimulation and voice detection task. At the end of the experiment, participants filled out the PDI questionnaire.

3.3.3.3.1 Sensorimotor blocks: Assessment of illusory effects

Identical to our previous studies (Blanke et al., 2014; Salomon et al., 2020), during sensorimotor blocks participants manipulated a robotic system that consists of two integrated units: the front part – a commercial haptic interface (Phantom Omni, SensAble Technologies) – and the back part – a three degree-of-freedom robot (Hara et al., 2011) (Figure 1). Blindfolded participants were seated between the front and back parts of the robot and were asked to perform repeated poking movements with their right index finger using the front part. Participants' pokes were replicated by the back part, thus applying corresponding touches on participants' backs. The touches were mediated by the robot either in synchronous (without delay) or in asynchronous (with 500 milliseconds delay) fashion, creating different degrees of sensorimotor conflict between the upper limb movement and somatosensory feedback on the back (Blanke et al., 2014; Faivre et al., 2020; Salomon et al., 2020).

Following a two-minute-long sensorimotor stimulation (both synchronous and asynchronous), participants filled out a short questionnaire. Specifically, on a Likert scale from 0 (not at all) to 6 (very strong), after each block, participants rated the strength of illusory self-touch ("I felt as if I was touching my back by myself"), somatic passivity ("I felt as if someone else was touching my back") and presence hallucination ("I felt as if someone was standing close to me"). Questionnaire contained an additional control item ("I felt as if I had three bodies."). The order of the two blocks (synchronous and asynchronous) was counterbalanced across participants.

3.3.3.3.2 Staircase blocks: Detecting hearing thresholds

Participants' individual hearing thresholds were estimated with a voice detection task combined with a one-up-one-down staircase procedure (Cornsweet, 1962). During the task, participants were continuously hearing short bursts of pink noise and were instructed to report whether they heard a voice in the noise by pressing on a button after the noise offset. Each burst of noise lasted for 3.5 seconds and voice onset randomly occurred in a period between 0.5 and 2.5 seconds after the noise onset, ensuring a minimum of 0.5 seconds of noise before and after the presentation of a voice recording. Following participants' response in each trial (i.e. a button click after the noise offset), an inter-trial interval jittered between 1 and 1.5 seconds.

The staircase procedure employed only other-voice stimuli and consisted of two blocks, one starting from a suprathresholded (top-down block) and another from subthresholded (bottom-up block) sound intensity level, counterbalanced across participants. In both staircase blocks, each word was presented four times in a randomized order, resulting in 36 trials. Threshold in each block was computed as a mean value from the last 15 trials and the average of the two thresholds was considered as participants' hearing threshold. No differences in detectability between self-voice and other-voice stimuli, as well as between different words were assured in a pilot study.

3.3.3.3.3 Task blocks: Combining a voice detection task with sensorimotor stimulation

During Task blocks, participants were performing the voice detection task while being exposed to sensorimotor stimulation (i.e. while they manipulated the robotic device). Task blocks differed based on the type of sensorimotor stimulation (synchronous, asynchronous), as well as of vocal stimuli (self, other). Thus, each participant completed four Task blocks (synchronous-self, synchronous-other, asynchronous-self, asynchronous-other) and had a unique order of blocks (i.e. we tested 24 participants to match 24 possible permutations of Task blocks). Task blocks started with 30 seconds of sensorimotor stimulation, followed by a concomitant voice detection task (Figure 33). Importantly, throughout the auditory task, participants continued manipulating the robot and auditory stimuli were not time-locked to participants' movements. The voice detection task was identical to the task in Staircase blocks, with the addition of 18 trials that contained only noise (i.e. no-voice trials). No-voice trials were randomized together with 45 trials containing a voice (i.e. each word was presented five times within a block), resulting in 63 trials per block. An adaptive staircase procedure was maintained throughout the block to ensure that the voices were presented at hearing threshold.
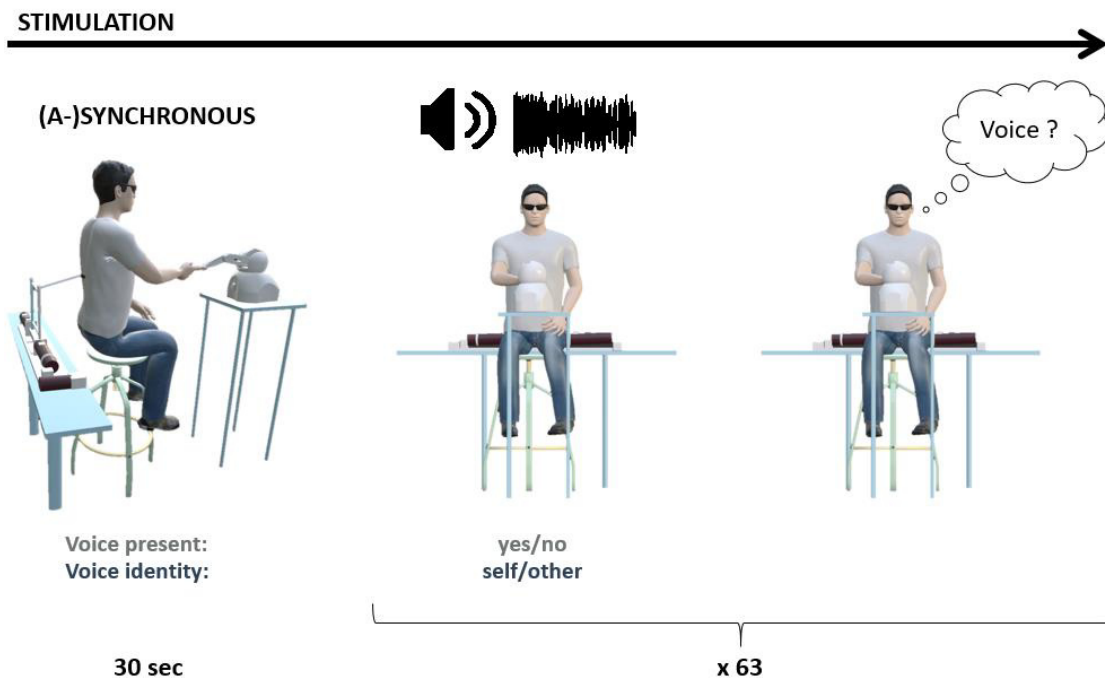
Figure 33. Task block design.

The block started with 30 seconds of sensorimotor stimulation, which was followed by simultaneous voice detection task. While manipulating the robotic device, participants were hearing bursts of noise and were instructed to report whether they heard a voice in the noise. Out of 63 trials, 45 contained a voice presented at hearing threshold. Within a block, the voices either belonged to participant (self) or to a stranger (other).

3.3.3.4 Statistical analysis

Statistical analysis and plotting was performed in R (R Core Team, 2020), using notably the lme4 (Bates, Mächler, Bolker, & Walker, 2015), lmerTest (Kuznetsova, Brockhoff, & Christensen, 2018), and afex (Singmann, Bolker, Westfall, & Aust, 2019) packages. The results were illustrated using sjplot (Lüdecke, 2018) and ggplot2 (Wickham, 2016) packages.

3.3.3.4.1 Vocal false alarms

Serving as a measure of experimentally-induced auditory hallucinations, our primary research interest was to identify the effects of sensorimotor stimulation on vocal false alarm rate. Thus, on no-voice trials, we conducted a mixed-effects binomial regression with Response as dependent variable and Stimulation (synchronous, asynchronous), Voice (self, other) and Gender (male, female) as fixed effects, and participants as random effect. The Response-variable indicates whether participants heard a voice in the noise, thus for the no-voice trials it represents the false alarm rate (whereas for the trials containing a voice in the noise, it stands for the hit rate). An interaction term was added between the effects of Stimulation and Voice. The Gender effect was added to the regression because in Study 1 there were more female participants. Random effects included a by-participant random intercept. By-participant random slopes for the main effects were added following model selection based on maximum likelihood. As a control for voice detectability, we also conducted equivalent analyses for the trials with voices present in noise.

3.3.3.4.2 Illusory effects

Ratings in questionnaire items were assessed by a mixed-effects linear regression containing a fixed effect of Stimulation (synchronous, asynchronous) and by-subject random intercepts. As the direction of the effect on each illusion is known from our previous work (Blanke et al., 2014; Salomon et al., 2020), we applied one-tailed tests.

3.3.3.4.3 Vocal false alarms and illusory effects

For the questionnaire items that significantly differed between the two sensorimotor stimulations (synchronous, asynchronous), we additionally explored whether the illusion assessed by the corresponding questionnaire item affected false alarm rate in the voice detection task. Specifically, to the mixed-effect binomial regression described above (with Response as a dependent variable) we added an additional fixed effect Illusion, with values represented as Likert-scale ratings (0-6) given for the corresponding questionnaire item and sensorimotor stimulation. The effect of Illusion was related with an interaction term with the effect of Condition. Similarly, we explored the effects of delusional ideation on false alarm rate, by adding PDI score as a covariate to the equivalent mixed-effect binomial regression, and forming a two-way interaction together with the effect of Condition.

3.3.4 Results

3.3.4.1 Vocal false alarms

In Study 1, a mixed-effects binomial regression revealed a main effect of Stimulation (estimate=-0.56, Z=-2.06, p=0.04), indicating a higher false alarm rate during asynchronous stimulation. We further observed a main effect of Gender (estimate=-2.4, Z=-1.99, p=0.046), revealing more false alarms in female participants. The effect of Voice was not significant (estimate=-0.36, Z=-1.3, p=0.193), however, it significantly interacted with the effect of Stimulation (estimate=0.97, Z=2.5, p=0.013). Further analysis of this interaction indicated that during the blocks containing other-voice stimuli, false alarm rate was increased with asynchronous stimulation (estimate=-0.52, Z=-1.95, p=0.051), whereas during self-voice blocks, false alarm rate increased with synchronous stimulation (estimate=0.57, Z=1.9, p=0.058) (Figure 34, left).
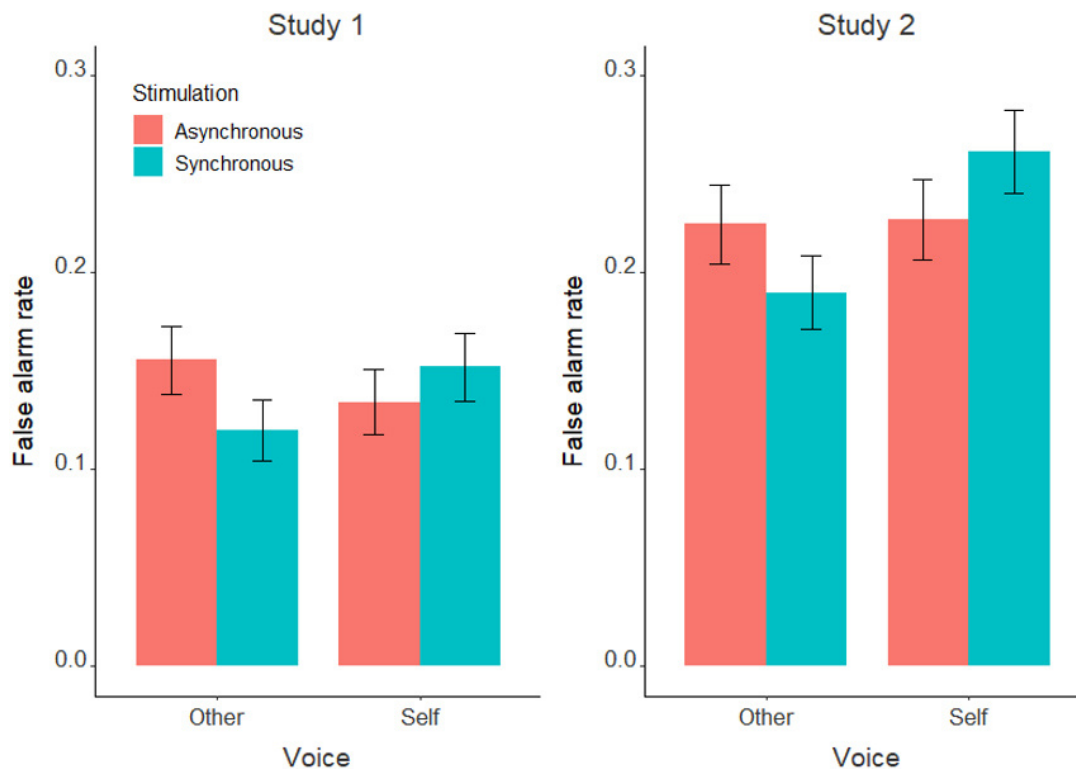
Figure 34. Vocal false alarms.

Vocal false alarm rates observed in Study 1 (left) and Study 2 (right). Height of bar plots indicates mean rate, and error bars 95% confidence intervals. In both studies, asynchronous stimulation increased false alarm rate in blocks containing other-voice stimuli, whereas synchronous stimulation increased false alarms in self-voice blocks.

In Study 2, we replicated the interaction between Stimulation and Voice (estimate=1.14, Z=2.97, p=0.003), and it revealed the same post-hoc effects – in other-voice blocks false alarms increased with asynchronous (estimate=-0.64, Z=-2.15, p=0.031), whereas in self-voice blocks with synchronous stimulation (estimate=0.53, Z=1.98, p=0.048) (Figure 34, right). Also, there were, again, more false alarms during asynchronous stimulation (estimate=-0.63, Z=-2.21, p=0.027) and no difference in false alarms between the two voices (estimate=0.03, Z=0.14, p=0.887). In Study 2, the effect of Gender was not significant (estimate=-1, Z=-0.71, p=0.479).

3.3.4.2 Illusory effects

In both experiments, sensorimotor stimulation induced the expected illusionary effects (Blanke et al., 2014; Salomon et al., 2020). Somatic passivity was rated higher during asynchronous stimulation (Study 1: estimate=-1.08, t(24)=-3.68, p=0.001; Study 2: estimate=-0.58, t(24)=-1.81, p=0.041), whereas Self-touch ratings were higher during synchronous compared to asynchronous stimulation (Study 1: estimate=0.79, t(24)=2.21, p=0.019; Study 2: estimate=0.83, t(24)=3.46, p=0.001). Presence hallucination was, as expected, experienced more during asynchronous stimulation (Study 1: estimate=-0.5, t(24)=-2.68, p=0.007; Study 2: estimate=-0.67, t(24)=-1.92, p=0.033). Control questionnaire item was unaffected by sensorimotor stimulation (Study 1: estimate=-0.04, t(24)=-0.58, p=0.566; Study 2: estimate=0.13, t(24)=1.39, p=0.176). An overview of illusory effects is given in Figure 35.
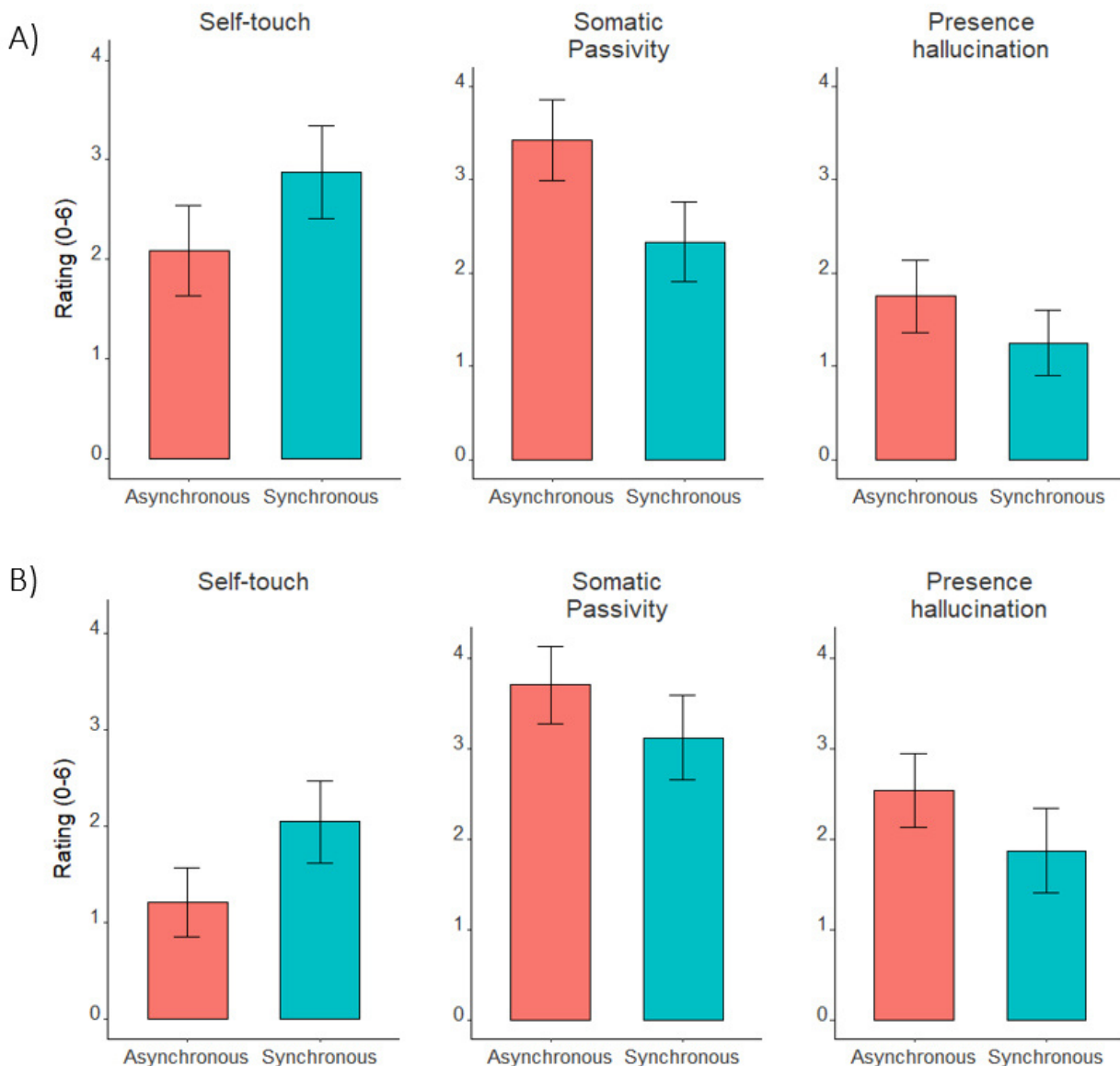
Figure 35. Illusory effects assessed in sensorimotor blocks in Study 1 (A) and Study 2 (B).
Height of bar plots indicates mean rating, and error bars 95% confidence intervals. In both studies, self-touch was higher during synchronous, whereas somatic passivity and presence hallucination during asynchronous stimulation.

3.3.4.3 Vocal false alarms and illusory effects

In Study 1, binomial mixed-effects analysis with the Response as dependent variable and Condition and Illusion as fixed effects, indicated a significant interaction between Somatic Passivity and Condition (estimate=-0.54, Z=-4.07, p<0.001). This effect, however, was not replicated in Study 2 (estimate=-0.04, Z=-0.35, p=0.724). Self-touch and Presence Hallucination did not have any effects on false alarms in either study (all p>0.05).

3.3.4.4 Delusional ideation

Contrary to the effects of robotically-induced illusions on vocal false alarms, the effect of delusional ideation was consistent in both studies. Thus, in both studies, we observed an interaction between Stimulation and PDI (Study 1: estimate=-0.17, Z=-2.1, p=0.036; Study 2: estimate=-0.1, Z=-1.9, p=0.058). Further investigation of these interactions revealed the main effects of PDI only for false alarms occurring during asynchronous stimulation (Study 1: estimate=0.32, Z=1.98, p=0.048; Study 2: estimate=0.32, Z=1.68, p=0.062). PDI was not related to false alarms that occurred during synchronous stimulation (Study 1: estimate=0.14, Z=0.69, p=0.494; Study 2: estimate=0.26, Z=1.23, p=0.22). Moreover, there was the main effect of PDI in both studies (Study 1: estimate=0.33, Z=1.96, p=0.049; Study 2: estimate=0.33, Z=2.15, p=0.032). Together, these effects indicate that the higher participants scored on delusional ideation inventory, the more false alarms they made during the auditory task and, moreover, that this increase was steeper during asynchronous stimulation (Figure 36).
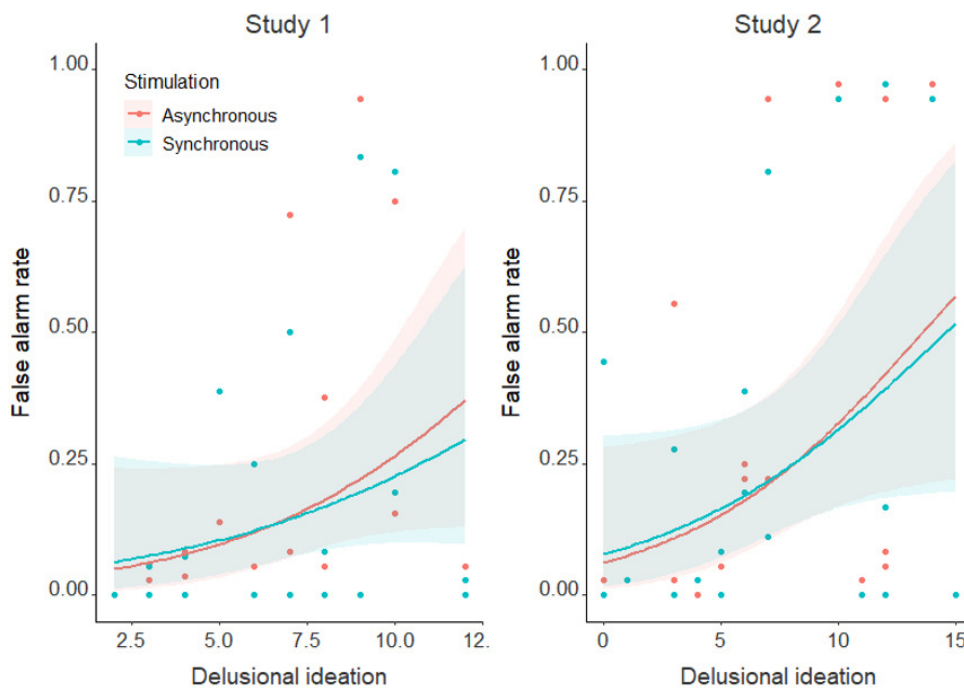
Figure 36. Delusional ideation and false alarms.
Increase in delusional ideation score was related to an increase in vocal false alarms rate in both studies. Shaded areas around each curve represent 95% confidence intervals.

3.3.4.5 Hit rate

There were no significant effects of sensorimotor stimulation and voice identity in both studies. In Study 1, binomial mixed-effects on yes-voice trials revealed a tendency for the main effect of Stimulation (estimate=-0.17, Z=-1.87, p=0.061), showing a higher hit rate during asynchronous stimulation. Hit rates were unaffected by Voice (estimate=-0.01, Z=-0.13, p=0.897) and Gender (estimate=-0.14, Z=-0.41, p=0.679). The interaction between Voice and Stimulation also indicated a tendency towards significance (estimate=0.24, Z=1.83, p=0.067). In Study 2, none of the borderlining effects proved significant. There was no main effect of Stimulation (estimate=-0.06, Z=-0.65, p=0.517), nor it interacted with Voice (estimate=0.12, Z=0.89, p=0.376). There were no significant effects of Gender (estimate=-0.71, Z=-1.74, p=0.081) nor Voice (estimate=-0.09, Z=-0.97, p=0.331).

Binomial mixed-effects regression with the dependent variable Response and fixed effects of PDI, Stimulation and Gender revealed no relationship between PDI and hit rate in either study. There were no main effects of PDI (Study 1: estimate=0.06, Z=1.02, p=0.31; Study 2: estimate=0.09, Z=1.7, p=0.09), nor it interacted with the effect of Stimulation (Study 1: estimate=-0.05, Z=-1.43, p=0.15; Study 2: estimate=0, Z=-0.23, p=0.82). Hit rate analysis is summarized in Figure 37.
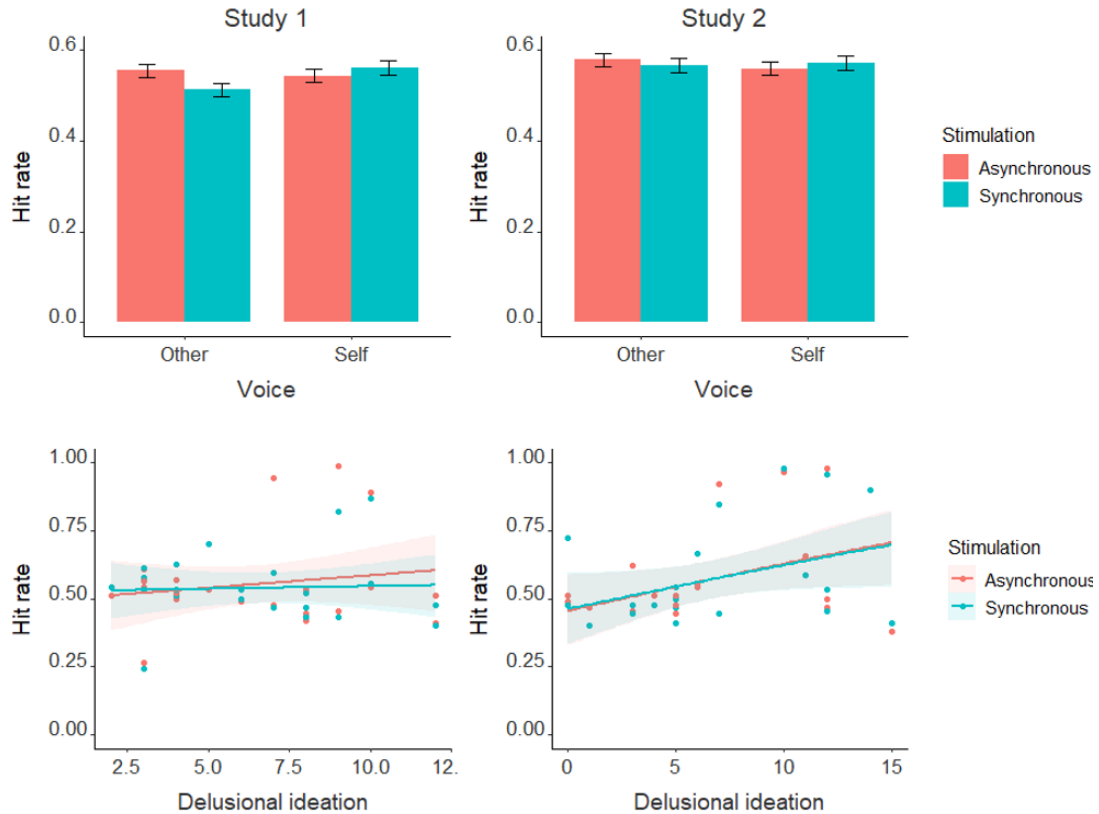
Figure 37. Hit rates.
Hit rates were not affected by experimental manipulation (top), nor were they related to delusional ideation score (bottom), in either study.

.

3.3.5   Discussion

In two independent cohorts of healthy participants, we observed that specific sensorimotor robotic stimulation could induce AVH, as indicated by an increase in the false alarm rate in a voice detection task. In addition, the rate of the vocal false percepts was positively related to delusional ideation. Those false vocal percepts were additionally modulated by the identity of voices to be detected. Specifically, sensorimotor stimulation associated with other-agency sensations (i.e. somatic passivity and presence hallucination) induced more false alarms in blocks with other-voice stimuli, whereas stimulation associated with illusory self-touch sensations increased false percepts in blocks containing self-voice stimuli.

Here, we demonstrated an experimental procedure able to causally induce AVH in healthy participants in a fully controlled environment. During asynchronous sensorimotor stimulation – that is related to mild bodily hallucinations (Blanke et al., 2014; Salomon et al., 2020) – participants reported more false alarms in a voice detection task, compared to the synchronous stimulation. The importance of this finding is threefold. First, this demonstrates a new experimental paradigm able to induce controlled auditory hallucinations in healthy, non-hallucinating individuals. Previous work employed conditioning paradigms, thereby experimentally manipulating visuo-auditory contingencies (Davies, Davies, & Bennett, 1982; Ellson, 1941; Powers et al., 2017), or using mere suggestion (Barber & Calverley, 1964; Seashore, 1895). Contrary to previous studies, our paradigm contained no trial-by-trial modulations, but instead showed a state-dependent effect on audition – while participants continuously perceived spatiotemporal conflicts, they were more prone to hearing non-existing voices in noise. Second, this demonstrates that sensorimotor processing exerts an orthogonal manipulation on voice perception, moreover without an explicit temporal relationship between sensorimotor stimulation and the onset of auditory stimuli. Namely, participants were moving freely and voice presentation was time-locked neither to

the movements nor the corresponding touches. In addition, inter-trial interval randomly jittered, making it impossible for participants to predict the onset of stimuli and thus adapt their movements. This extends our previous results linking sensorimotor conflicts to alterations in voice loudness perception (Orepic et al., 2020) to voice detection. Hence, the third contribution of this finding is that it consolidates the self-monitoring account for AVH, albeit not directly through prediction mechanisms related to voice production. Namely, compared to the current viewpoint (Ford, Roach, Faustman, & Mathalon, 2007; Frith, 1992; Frith & Done, 1989; Moseley, Fernyhough, & Ellison, 2013; Shergill et al., 2003), hallucinations arose not as an impairment in speech feedforward mechanisms, but as an impairment in somatosensory, bodily self-monitoring process. Importantly, sensorimotor stimulation did not affect hit rate in voice detection task, controlling for differences in auditory detectability between experimental stimuli (self and other voice) and sensorimotor stimulations (synchronous and asynchronous) and further excluding attentional confounds.

Specific sensorimotor stimulation had a different effect on false vocal percepts depending on the identity of heard voices. As predicted, during asynchronous stimulation there were more false percepts in other-voice blocks. However, to our surprise, false alarms were also increased during synchronous stimulation, albeit in self-voice blocks. Asynchronous stimulation (as has also been replicated in both studies here) induces a presence hallucination – a subjective experience related to an alien agent – that is by definition an entity distinct from the self. If one perceives a presence of someone else, it is to be expected that one also hears a voice of someone else. However, why would one hear her own voice dependent on sensorimotor stimulation? The answer might lie in illusory sensations of agency. Namely, synchronous stimulation facilitates illusory self-touch sensations – participants have the impression of being the one touching their own back, despite the spatial conflict to the performed movement (i.e. it is impossible to touch own back by doing a forward poking movement). Those sensations might be interpreted as a self-attribution bias (Farrer, Franck, Paillard, & Jeannerod, 2003; Hauser et al., 2011; Tsakiris, Haggard, Franck, Mainy, & Sirigu, 2005), occurring due to temporal congruence of spatially conflicting stimuli. Similar effects are thought to account for rubber-hand (Botvinick & Cohen, 1998) and full-body (Lenggenhager, Tadi, Metzinger, & Blanke, 2007) illusions, where temporal congruence of spatially-conflicting visuotactile stimulation leads to a self-attribution of the rubber hand or an avatar, respectively. Similarly, somatic passivity that is reported in asynchronous stimulation, might constitute an other-attribution bias, due to temporal incongruence between the movement and the resulting touch. We propose that these self- and other-attribution biases might be reflected in the vocal false alarms of the corresponding identity.

How could the relationship between attribution biases and false alarms be instantiated? In both studies, we observed a positive relationship between false alarms and delusional ideation, which was more pronounced during asynchronous stimulation. Delusions are unfounded yet pertinacious beliefs that can be found in varying degrees also in the general population (Peters et al., 2004). In the study of (Teufel et al., 2010), delusional ideation was related to self-monitoring impairments (Shergill, Bays, Frith, & Wolpert, 2003). Specifically, individuals with higher PDI scores exhibited less self-attenuation – resembling the behavioral patterns of schizophrenia patients (Shergill et al., 2005). Our data is in concordance with this finding, as we observed a stronger relationship between PDI and false alarms during asynchronous, compared to synchronous stimulation, that is characterized by stronger sensorimotor conflicts, as well as decreases in self-attenuation (Orepic et al., 2020). Thus, we replicated a relationship between delusion proneness and the degree of self-monitoring impairments and additionally associated delusion proneness to induced false vocal percepts.

It has further been shown that individuals with high delusion proneness – as measured again by higher PDI scores – rely more on prior expectations than on sensory evidence while performing specifically tailored behavioral tasks (Schmack et al., 2013; Teufel et al., 2015). This suggests that PDI might constitute a measure of overly strong perceptual priors that impose top-down effects on perception (Adams, Brown, & Friston, 2015; Fletcher & Frith, 2009; Sterzer et al., 2018). Thus, a positive relationship between increases in delusional ideation and false alarms in our studies might indicate an effect driven by perceptual priors attributed with undue precision (Corlett et al., 2019). We speculate that repeatedly hearing a voice of specific identity (self or other) throughout our experimental blocks, in fact, creates an expectation about the identity of the voices to follow – i.e. if one continuously hears consecutive self-voice stimuli, one might expect to hear the same type of stimulus again in the near future. Crucially, the directionality of the imposed auditory prior (self or other) might hence interfere with the attribution biases characterizing the concomitant sensorimotor stimulation, moreover in a complementary fashion. Thus, self-voice prior combined with self-attribution bias might lead to an increase in self-voice false alarms, and vice versa.

In a recent theoretical piece, (Leptourgos & Corlett, 2020) proposed a conceptual model that fits well with this interpretation. They suggest there might be two distinct hierarchies at play – one facilitating self-monitoring, and the other agency-related priors – whereby both make separate inferences from the same sensory input, which are combined according to a weighted cue combination mechanism into a final construct portraying sense of agency. According to this model, self-monitoring deficits resulting from erroneous feedforward sensorimotor processing result in a noisy sensory input that, in turn, elicits compensatory alterations in

precision-weighting of existing self- and other-related priors (Corlett et al., 2019), attributing the noisy input to the most plausible agent. Applied to our data, robotically-mediated sensorimotor conflicts might create noisy and unpredictable sensory inputs (touches on the back), which are compensated by an increase in precision of high-level priors (self or other), explaining away the noisy sensory input (e.g. self-attribution bias during synchronous and other-attribution bias during asynchronous stimulation). The bidirectionality is reflected in the concomitant 'auditory prior' – i.e. the expectation to hear a voice following repeated exposure to that voice – such that other-prior makes it more likely to hear other-voice, and vice versa. However, it should be noted that this is pure speculation and future experiments should be specifically designed to address individual parameters of such a model, exploring its applicability to this paradigm.

In conclusion, here we demonstrated a sensorimotor-robotic procedure and method able to induce AVH in healthy individuals and in a fully controlled laboratory environment. Specifically, we showed that conflicting sensorimotor stimulation can selectively induce vocal false percepts – whereby spatial sensorimotor conflict facilitates self-, and spatiotemporal conflicts other-voice false percepts – an effect related with delusion proneness. Besides the novelty and the important methodological impact, these results shed new light on AVH phenomenology, representing experimental support for both prominent albeit seemingly opposing accounts – portraying AVH as a hybrid between deficits in self-monitoring and hyper-precise priors.

Author contribution

3.3.6 References

Adams, R. A., Brown, H. R., & Friston, K. J. (2015). Bayesian inference, predictive coding and delusions. Avant, V(3), 51–88. https://doi.org/10.12849/50302014.0112.0004

Barber, T. X., & Calverley, D. S. (1964). An experimental study of "hypnotic" (auditory and visual) hallucinations. The Journal of Abnormal and Social Psychology, 68(1), 13.

Barkus, E., Smallman, R., Royle, N., Barkus, C., Lewis, S., & Rushe, T. (2011). Auditory false perceptions are mediated by psychosis risk factors. Cognitive Neuropsychiatry, 16(4), 289–302. https://doi.org/10.1080/13546805.2010.530472

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software. https://doi.org/10.18637/jss.v067.i01

Bauer, S. M., Schanda, H., Karakula, H., Olajossy-Hilkesberger, L., Rudaleviciene, P., Okribelashvili, N., … Stompe, T. (2011). Culture and the prevalence of hallucinations in schizophrenia. Comprehensive Psychiatry, 52(3), 319–325. https://doi.org/10.1016/j.comppsych.2010.06.008

Bernasconi, F., Blondiaux, E., Potheegadoo, J., Stripeikyte, G., Bejr-kasem, H., Pagonabarraga, J., … Blanke, O. (2020). Sensorimotor hallucinations in Parkinson's disease. BioRxiv.

Blakemore, S. J., Smith, J., Steel, R., Johnstone, E. C., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. Psychological Medicine, 30(5), 1131–1139. https://doi.org/10.1017/S0033291799002676

Blanke, O., Pozeg, P., Hara, M., Heydrich, L., Serino, A., Yamamoto, A., … Rognini, G. (2014). Neurological and robot-controlled induction of an apparition. Current Biology, 24(22), 2681–2686. https://doi.org/10.1016/j.cub.2014.09.049

Botvinick, M., & Cohen, J. D. (1998). Rubber hand 'feels' what eyes see. Nature, 391(February), 756.

Brainard, D. H. (1997). The Psychophysics Toolbox. Spatial Vision, 10(4), 433–436. https://doi.org/10.1163/156856897X00357

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-day, B., Schmack, K., & Iii, A. R. P. (2019). Hallucinations and Strong Priors. Trends in Cognitive Sciences, 23(2), 114–127. https://doi.org/10.1016/j.tics.2018.12.001

Cornsweet, T. N. (1962). The Staircase-Method in Psychophysics. The American Journal of Psychology, 75(3), 485–491.

Davies, P., Davies, G. L., & Bennett, S. (1982). An effective paradigm for conditioning visual perception in human subjects. Perception, 11(6), 663–669.

Ellson, D. G. (1941). Hallucinations produced by sensory conditioning. Journal of Experimental Psychology 28, No. 1 (1941): 1., 28(1), 1–20.

Faivre, N., Vuillaume, L., Bernasconi, F., Salomon, R., Blanke, O., & Cleeremans, A. (2020). Sensorimotor conflicts alter metacognitive and action monitoring. Cortex, 124(22), 2681–2686. https://doi.org/10.1016/j.cortex.2019.12.001

Farrer, C., Franck, N., Paillard, J., & Jeannerod, M. (2003). The role of proprioception in action recognition. Consciousness and Cognition, 12(4), 609–619. https://doi.org/10.1016/S1053-8100(03)00047-3

Feinberg, I. (1978). Efference copy and corollary discharge: implications for thinking and its disorders. Schizophrenia Bulletin, 4(4), 636–640. https://doi.org/10.1093/schbul/4.4.636

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nature Reviews Neuroscience, 10(1), 48–58. https://doi.org/10.1038/nrn2536

Ford, J. M., Mathalon, D. H., Whitfield, S., Faustman, W. O., & Roth, W. T. (2002). Reduced communication between frontal and temporal lobes during talking in schizophrenia. Biological Psychiatry, 51(6), 485–492. https://doi.org/10.1016/S0006-3223(01)01335-X

Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2007). Synch before you speak: Auditory hallucinations in schizophrenia. American Journal of Psychiatry, 164(3), 458–466. https://doi.org/10.1176/appi.ajp.164.3.458

Frith, C. D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. Psychological Medicine, 17(3), 631–648. https://doi.org/10.1017/S0033291700025873

Frith, C. D. (1992). The cognitive neuropsychology ofschizophrenia. Erlbaum. https://doi.org/10.1016/0028-3932(93)90120-O

Frith, Chris D., Blakemore, S. J., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. Brain Research Reviews, 31(2–3), 357–363. https://doi.org/10.1016/S0165-0173(99)00052-1

Frith, Christopher D., & Done, D. J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. Psychological Medicine, 19(2), 359-363. https://doi.org/10.4324/9781315630502

Gould, L. N. (1948). Verbal hallucinations and activity of vocal musculature; an electromyographic study. The American Journal of Psychiatry, 105, 367–72. https://doi.org/10.1176/ajp.105.5.367

Green, M. F., & Kinsbourne, M. (1990). Subvocal activity and auditory hallucinations: Clues for behavioral treatments? Schizophrenia Bulletin, 16(4), 617–625.

Green, P., & Preston, M. (1981). Reinforcement of vocal correlates of auditory hallucinations by auditory feedback: a case study. The British Journal of Psychiatry, 139(3), 204–208.

Hara, M., Rognini, G., Evans, N., Blanke, O., Yamamoto, A., Bleuler, H., & Higuchi, T. (2011). A novel approach to the manipulation of body-parts ownership using a bilateral master-slave system. In IEEE International Conference on Intelligent Robots and Systems (pp. 4664–4669). https://doi.org/10.1109/IROS.2011.6048519

Harkavy-Friedman, J. M., Kimhy, D., Nelson, E. A., Venarde, D. F., Malaspina, D., & Mann, J. J. (2003). Suicide attempts in schizophrenia: the role of command auditory hallucinations for suicide. The Journal of Clinical Psychiatry., 64(8), 871–874.

Hauser, M., Knoblich, G., Repp, B. H., Lautenschlager, M., Gallinat, J., Heinz, A., & Voss, M. (2011). Altered sense of agency in schizophrenia and the putative psychotic prodrome. Psychiatry Research, 186(2–3), 170–176. https://doi.org/10.1016/j.psychres.2010.08.003

Hoffman, R. E., Fernandez, T., Pittman, B., & Hampson, M. (2011). Elevated functional connectivity along a corticostriatal loop and the mechanism of auditory/verbal hallucinations in patients with schizophrenia. Biological Psychiatry, 69(5), 407–414.

Jacobson, E. (1931). Electrical measurements of neuromuscular states during mental activities: V. Variation of specific muscles contracting during imagination. American Journal of Physiology-Legacy Content, 96(1), 115–121. https://doi.org/10.1152/ajplegacy.1931.96.1.115

Jalenques, I., Enjolras, J., & Izaute, M. (2013). Valence émotionnelle des mots. L'Encéphale, 39, 189–197.

Kleiner, M., Brainard, D. H., Pelli, D. G., Ingling, A., Murray, R., & Broussard, A Ingling, R Murray, C. (2007). What's new in Psychtoolbox-3? Perception, 36(14), 1–16. https://doi.org/10.1068/v070821

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lehman, A. F., Lieberman, J. A., Dixon, L. B., McGlashan, T. H., Miller, A. L., Perkins, D. O., … Cook, I. (2004). Practice guideline for the treatment of partients with schizophrenia. American Journal of Psychiatry, 161(2 SUPPL), 1–56.

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video Ergo Sum: Manipulating Bodily Self-Consciousness. Science, 317(5841), 1096–1099. https://doi.org/10.1126/science.1143439

Leptourgos, P., & Corlett, P. R. (2020). Embodied Predictions, Agency, and Psychosis. Frontiers in Big Data, 3(August), 27. https://doi.org/10.3389/FDATA.2020.00027

Livesay, J., Liebke, A., Samaras, M., & Stanley, A. (1996). Covert speech behavior during a silent language recitation task. Perceptual and Motor Skills, 83(3_suppl), 1355–1362.

Lüdecke, D. (2018). sjPlot: Data Visualization for Statistics in Social Science. R package version 2.6.2. https://doi.org/10.5281/zenodo.1308157

McCarthy-Jones, S., Trauer, T., MacKinnon, A., Sims, E., Thomas, N., & Copolov, D. L. (2014). A new phenomenological survey of auditory hallucinations: Evidence for subtypes and implications for theory and practice. Schizophrenia Bulletin, 40(1), 225–235. https://doi.org/10.1093/schbul/sbs156

McGuigan, F. J. (1966). Covert oral behavior and auditory hallucinations. Psychophysiology, 3, 73–80. https://doi.org/10.1111/j.1469-8986.1966.tb02682.x

McGuigan, F. J., & Dollins, A. B. (1989). Patterns of covert speech behavior and phonetic coding. The Pavlovian Journal of Biological Science, 24, 19–26. https://doi.org/10.1007/BF02964527

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. Neural Networks, 9(8), 1265–1279.

Moseley, P., Fernyhough, C., & Ellison, A. (2013). Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. Neuroscience and Biobehavioral Reviews, 37(10), 2794–2805. https://doi.org/10.1016/j.neubiorev.2013.10.001

Moseley, P., Fernyhough, C., & Ellison, A. (2014). The role of the superior temporal lobe in auditory false perceptions: A transcranial direct current stimulation study. Neuropsychologia, 62(1), 202–208. https://doi.org/10.1016/j.neuropsychologia.2014.07.032

Orepic, P., Park, H. D., Rognini, G., Faivre, N., & Blanke, O. (2020). Breathing affects self-other voice discrimination in a bodily state with increased otherness. PsyArXiv. https://doi.org/10.31234/osf.io/t9saq

Orepic, Pavo, Rognini, G., Kannape, O. A., Faivre, N., & Blanke, O. (2020). Sensorimotor conflicts induce somatic passivity and louden quiet voices in healthy listeners. BioRxiv. https://doi.org/https://doi.org/10.1101/2020.03.26.005843

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10(4), 437–442. https://doi.org/10.1163/156856897X00366

Peters, E., Joseph, S., Day, S., & Garety, P. (2004). Measuring delusional ideation: the 21-item Peters et al. Delusions Inventory (PDI). Schizophrenia Bulletin, 30(4), 1005–1022.

Powers, A R, Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. Science, 357(August), 596–600. https://doi.org/10.1126/science.aan3458

Powers, Albert R., Kelley, M. S., & Corlett, P. R. (2017). Varieties of voice-hearing: Psychics and the psychosis continuum. Schizophrenia Bulletin, 43(1), 84–98. https://doi.org/10.1093/schbul/sbw133

R Core Team. (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Salomon, R., Progin, P., Griffa, A., Rognini, G., Do, K. Q., Conus, P., … Blanke, O. (2020). Sensorimotor Induction of Auditory Misattribution in Early Psychosis. Schizophrenia Bulletin. https://doi.org/10.1093/schbul/sbz136

Schmack, K., de Castro, A. G. C., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J. D., … Sterzer, P. (2013). Delusions and the role of beliefs in perceptual inference. Journal of Neuroscience, 33(34), 13701–13712. https://doi.org/10.1523/JNEUROSCI.1778-13.2013

Seashore, C. E. (1895). Measurements of Illusions and Hallucinations in normal life. Reprinted from Studies from the Yale Psychological Laboratory, Vol. 3.

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. Annual Review of Neuroscience, 33, 89–108. https://doi.org/10.1146/annurev-neuro-060909-153135

Shergill, S. S., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2003). Two Eyes for an Eye: The Neuroscience of Force Escalation. Science, 301(5630), 187–187. https://doi.org/10.1126/science.1085327

Shergill, S. S., Brammer, M. J., Fukuda, R., Williams, S. C. R., Murray, R. M., & Mcguire, P. K. (2003). Engagement of brain areas implicated in processing inner speech in people with auditory hallucinations. He British Journal of Psychiatry, 182(6), 525–531. https://doi.org/10.1192/02-505

Shergill, S. S., Samson, G., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2005). Evidence for sensory prediction deficits in schizophrenia. American Journal of Psychiatry, 162(12), 2384–2386. https://doi.org/10.1176/appi.ajp.162.12.2384

Shergill, S. S., White, T. P., Joyce, D. W., Bays, P. M., Wolpert, D. M., & Frith, C. D. (2014). Functional magnetic resonance imaging of impaired sensory prediction in schizophrenia. JAMA Psychiatry, 71(1), 28–35. https://doi.org/10.1001/jamapsychiatry.2013.2974

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2019). afex: Analysis of Factorial Experiments. R package version 0.23-0. https://cran.r-project.org/package=afex. Retrieved from https://cran.r-project.org/package=afex

Sommer, I. E., Daalman, K., Rietkerk, T., Diederen, K. M., Bakker, S., Wijkstra, J., & Boks, M. P. M. (2010). Healthy individuals with auditory verbal hallucinations; Who are they? psychiatric assessments of a selected sample of 103 subjects. Schizophrenia Bulletin, 36(3), 633–641. https://doi.org/10.1093/schbul/sbn130

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., … Corlett, P. R. (2018). The Predictive Coding Account of Psychosis. Biological Psychiatry, 84(9), 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015

Teufel, C., Kingdon, A., Ingram, J. N., Wolpert, D. M., & Fletcher, P. C. (2010). Deficits in sensory prediction are related to delusional ideation in healthy individuals. Neuropsychologia, 48(14), 4169–4172. https://doi.org/10.1016/j.neuropsychologia.2010.10.024

Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., … Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proceedings of the National Academy of Sciences of the United States of America, 112(43), 13401–13406. https://doi.org/10.1073/pnas.1503916112

Tsakiris, M., Haggard, P., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. Cognition, 96(3), 215–231. https://doi.org/10.1016/j.cognition.2004.08.002

Whitford, T. J. (2019). Speaking-Induced Suppression of the Auditory Cortex in Humans and Its Relevance to Schizophrenia. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 4(9), 791–804. https://doi.org/10.1016/j.bpsc.2019.05.011

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., & Grodd, W. (1996). Functional lateralization of speech production at primary motor cortex: A fMRI study. NeuroReport, 7, 2791–6. https://doi.org/10.1097/00001756-199611040-00077

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. Science, 269(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Woods, A., Jones, N., Alderson-Day, B., Callard, F., & Fernyhough, C. (2015). Experiences of hearing voices: Analysis of a novel phenomenological survey. The Lancet Psychiatry, 2(4), 323–331. https://doi.org/10.1016/S2215-0366(15)00006-1

General discussion

Motivated by the self-monitoring account of AVH, the present thesis revolved around self-other voice discrimination (SOVD) and experimental attempts of its perturbation with a robotic procedure able to induce impairments in bodily self-monitoring. It was thus divided into two conceptual parts and consisted of five studies. In this chapter, I will summarize all the findings and discuss them more broadly, thereby relating the findings of different studies with each other.

4.1 Self-voice perception

The first part of my thesis work investigated the nature of self-voice and made a case for self-voice being essentially a multimodal construct. The multimodal aspect was investigated by contrasting traditional presentation of self-voice stimuli through air conduction to the presentation of self-voices through bone conduction. In three different experiments, I observed a better performance in a sensitive SOVD task through bone, compared to air conduction. I also observed a higher increase in skin conductance response (SCR) to bone-conducted self-voices, indicating that even autonomic system differentiates the two types of presentations of the same self-voice recordings. Crucially, this advantage was not observed for familiar-other voice discrimination (FOVD), indicating an intimate relationship between bone conduction and self-voice, but not familiar-voice stimuli. I further observed that self-voice is more confused to a familiar voice, compared to an unfamiliar one, regardless of the acoustic similarity to those voices, suggesting that to a certain extent, familiarity processing is nevertheless involved while recognizing self-voice. Acoustic dissimilarity, however, correlated to SOVD performance, suggesting that it also plays a role in this process. This relationship was, however, observed only for male participants. Finally, I measured high-density EEG during the same task and indicated a topographic map specific to self-voice stimuli that projected to traditionally self-related areas, such as insulae and medial temporal regions, and, importantly, correlated with SOVD behavioral performance.

4.1.1 Multisensory nature of self-voice

Multisensory integration of bodily signals plays a crucial role in the construction of a coherent representation of the self (Blanke & Metzinger, 2009; Blanke et al., 2015; Gallagher, 2000; Jeannerod, 2003). This is corroborated by various reports of experimentally-induced spatial and temporal incongruence of bodily signals leading to altered states of bodily self-consciousness (Botvinick & Cohen, 1998; Braun et al., 2018; Kannape, Smith, Moseley, Roy, & Lenggenhager, 2019; Lenggenhager et al., 2007). Being an integral part of our self, self-voice is never perceived unimodally – namely, in a natural scenario of speaking, the sound of our voice is accompanied by vestibular and somatosensory afferents (Emami et al., 2012; Tremblay et al., 2003). Moreover, vestibular hearing happens to coincide with the range of our voice pitch (Todd et al., 2000) and somatosensory-auditory integration has already been functionally localized (Foxe et al., 2001; Murray et al., 2005). In an interesting study involving a robotic device that applies facial skin stretches, (Ito et al., 2009) showed that somatosensory stimulation alone can affect auditory perception of concomitant sounds. Crucially, speaking-associated vibrotactile excitation results from skull movements induced by bone conduction of our voice (Stenfelt, 2011). Thus, the multisensory nature of self-voice originates from bone conduction.

Previous self-voice studies consistently reported lower accuracy rates to self-voice compared to other-voice stimuli (Allen et al., 2005; Gur & Sackeim, 1979; Hughes & Nicholson, 2010; Rosa et al., 2008; Schuerman et al., 2015; Shuster, 1998) and attributed this discrepancy to the lack of bone conduction, albeit to the lack of physical transformation to the sound of self-voice imposed by bone conduction. Some tried to overcome this discrepancy by assessing subjective familiarity to voice-recordings as a function of different filters that were designed to mimic the transfer function of bone-conducted speech (Kimura & Yotsumoto, 2018; Maurer & Landis, 1990; Shuster & Durrant, 2003; Vurma, 2014; Won, Berger, & Slaney, 2014). As the exact transfer function is still unknown (Stenfelt, 2016), these studies yielded inconsistent or inconclusive results. Crucially, all the previous work has overseen the multimodal excitation that is missing in air-conducted self-voice recordings, and focused only on elucidating physical transformations. The data presented in this thesis suggests that the bone-conduction advantage for self-voice recognition mainly relies on multimodal excitation, and less on the imposed physical transformation to the sound of our voice. This was inferred upon the fact that there were no differences in FOVD between air and bone conduction – i.e., if physical transformation imposed by bone conduction were advantageous for SOVD, it could be expected to be disadvantageous to FOVD. Namely, the same way how such transformation could render self-voice more familiar, it should make familiar voice less familiar.

The work of my thesis shows that self-voice recognition can be increased with multimodal presentation of self-voice stimuli through a commercial bone conduction headset, i.e. even without speaking. As such headsets have become widely available, this finding might have a big impact on subsequent self-voice research.

4.1.2 Bone conduction effects on SCR and EEG

In addition to the performance in the SOVD task, bone-conduction advantage was observed in two other implicit measurements – evoked SCR and EEG responses. SCR was increased for self-voice compared to other-voice stimuli presented through bone conduction, with no differences in air conduction. SCR increase to self-voice stimuli has been long observed and associated to affective reactions accompanying hearing our voice in a voice recording (Douglas & Gibbins, 1983; Gur & Sackeim, 1979; Holzman, Rousey, & Snyder, 1966; Olivos, 1967). Interestingly, in those studies, SCR increased to self-voice stimuli even they were misattributed to another voice, suggesting that autonomic system is able to recognize self-voice stimuli at an early stage preceding conscious awareness. SCR increase has been observed also for self-face (Ameller et al., 2015; Sackeim & Gur, 1978; Sugiura et al., 2000) and extended to familiar voice (Douglas & Gibbins, 1983) and familiar face (Tranel & Damasio, 1985). SCR increase to self (or familiar) stimuli might thus serve as an implicit measure of self-recognition (or just recognition).

Contrary to previous studies, we did not observe a difference between self- and other voices with air conduction. We believe the reason for this lies in the task design. Namely, the task was to detect a voice in noise that was always significantly above hearing threshold, rendering the task very easy and monotonous. Most participants, who exhibited any SCR to vocal stimuli, did so only in the beginning of experimental blocks. This could be observed in a significant negative relationship between the increase in trial and SCR, showing that SCR decreased with the increase of experimental trials within a block. Monotony of the design might have led to a lack of evoked response. However, the fact that we did observe an increase in bone conduction suggests that bone conduction sensitizes such an implicit measure and thus might facilitate self-voice recognition even at the autonomic level. This potentially reflects an effect of multisensory integration occurring at an earlier stage compared to conscious recognition. Interestingly, in schizophrenia, the SCR increase to self/familiar voices seems to disappear (Ameller et al., 2015, 2017). It would thus be interesting to measure SCR of a similar task during asynchronous robotic stimulation, e.g., and relate SCR increase to the degree of impairments in self-monitoring, multisensory integration or some related mechanisms.

Similarly, the self-voice-associated network exhibited differential activation depending on sound conduction type. It was activated less with bone compared to air conduction. As the occurrence of this network negatively correlated to SOVD performance, we speculate that lower occurrence in bone conduction might as well reflect a better task performance. Regardless of the interpretation, we observed that a network related to self-voice processing was modulated by bone conduction, providing yet another piece of evidence associating bone conduction to self-voice.

4.1.3 Contributions of acoustical and familiarity processing

The data presented in this thesis shows that although self-voice is different from familiar voice processing (e.g. no differences in FOVD related to sound conduction and previous exposure), there are still some familiarity processes at play. In self-recognition task, that employed no voice morphing, the inability to recognize own voice (miss rate) was correlated with familiar-to-self and not with unfamiliar-to-self misattributions (false alarm rate). Thus, self voice was confused only with a familiar voice, regardless of acoustic similarity between the two voices. Acoustic similarity was assessed by correlating task performance with differences in vocal parameters that were previously shown to account for distinctiveness of (unfamiliar) voices (Baumann & Belin, 2010). We found a positive correlation only for male voices. Possible interpretations of this observation were discussed in the chapter belonging to Study 1, and they were purely speculative. Importantly, I seem to have found evidence suggesting that SOVD is a hybrid of low-level acoustic and high-level familiarity mechanisms.

4.2 Robotically-induced self-voice misperceptions

The second part contained different psychophysical assessments of self-voice perception during an orthogonal robotic-sensorimotor stimulation, investigating whether robotically-induced impairments in bodily self-monitoring could be extended to the auditory-verbal domain. I observed that sensorimotor stimulation able to induce somatic passivity and presence hallucination can not induce explicit alterations in SOVD, but can affect voice loudness perception in a manner that resembles self-attenuation deficits observed in schizophrenia. Explicit SOVD alterations were, however, shown to be modulated by breathing. Finally, when

adapting an implicit self-voice-related task – voice detection in noise – we observed an effect of robotic stimulation that was dependent on the identity of vocal stimuli (self vs other). Specifically, participants reported hearing more non-existing voices in the blocks containing other-voice stimuli while exposed to sensorimotor stimulation associated with other-agency sensations, and vice versa.

4.2.1 Impairments in bodily self-monitoring

Robotic procedure designed in our laboratory facilitates impairments in bodily-self monitoring by creating two types of sensorimotor stimulation – a synchronous and an asynchronous one. Synchronous stimulation consists of a spatial sensorimotor conflict between poking movements oriented to the front of the body and the corresponding touches on the back. Asynchronous stimulation contains an additional delay between the movement and somatosensory feedback, thus constituting spatiotemporal conflict. Temporal conflicts have been shown to cause a loss of agency, by manipulating sensory action consequences of upper-limb movements and related losses of hand movement agency (David, Newen, & Vogeley, 2008; Farrer, Bouchereau, Jeannerod, & Franck, 2008; Haggard, Clark, & Kalogeras, 2002; MacDonald & Paus, 2003; Sato & Yasuda, 2005; Stetson, Cui, Montague, & Eagleman, 2006; Tsakiris, Haggard, Franck, Mainy, & Sirigu, 2005). Crucially, when extending such manipulations to a torso-centered bodily system (Blanke & Metzinger, 2009; Park & Blanke, 2019), more than just agency loss, other-agency sensations can be introduced (Blanke et al., 2014; Salomon et al., 2020), together with a state of an altered bodily self-consciousness, including the alien agent (Blanke et al., 2014; Salomon et al., 2020). So far, we have assessed these changes by means of questionnaires – asking participants to rate the strength of illusory sensations after two minutes of sensorimotor stimulation in each condition. The strength of other-related agency sensations (i.e. somatic passivity and presence hallucination) was assessed as a difference in ratings following asynchronous and synchronous conditions, thereby accentuating the importance of the temporal conflict. Other researchers in my laboratory have investigated neural mechanisms underlying robotically-induced presence hallucination (Blondiaux, 2020) and related them to presence hallucinations experienced in Parkinson's disease (Bernasconi et al., 2020).

4.2.2 Change in loudness perception of voices

In Study 3, I reported that asynchronous sensorimotor stimulation induced a cross-modal effect on voice perception characterized by loudening of quiet voices. This effect was replicated in a follow-up study, which additionally contained two control conditions that involved no sensorimotor integration – in one, participants only moved the front part of the robot (motor-only), without being touched on their back, whereas in the other, they only received touches on the back (touch-only), without performing any movements. Quiet voices were again amplified in asynchronous condition and, importantly, there were no differences in loudness perception between synchronous and the two control conditions. Thus, the asynchronous condition, constituting an additional temporal conflict, led to alterations in voice loudness perception. Additionally, this effect was present only in participants who did not report other-agency sensations in the form of somatic passivity.

We proposed that the loudening effect might constitute a lack of sensory attenuation that is known to accompany sensorimotor conflicts (Kilteni & Ehrsson, 2017; Sato, 2008; Shergill et al., 2003; Stenner et al., 2014; Teufel, Kingdon, Ingram, Wolpert, & Fletcher, 2010; Weiss, Herwig, & Schütz-Bosbach, 2011a, 2011b). However, contrary to previous reports of (reduced) self-attenuation, in our design auditory stimuli were not time-locked to the movements, excluding traditional trial-by-trial effects on perception (e.g. a button press associated with a tone), suggesting rather a state-driven effect – i.e. while experiencing a specific sensorimotor stimulation, loudness perception is altered. The lack of sensory attenuation in hallucinating participants was reminiscent of a lack of sensory attenuation in schizophrenia patients (Blakemore et al., 2000; Ford, Gray, Faustman, Roach, & Mathalon, 2007; Ford et al., 2001; Shergill et al., 2005) as well as in healthy individuals depending on hallucination proneness (Asai, 2016; Teufel et al., 2010; Whitford, Mitchell, & Mannion, 2017). Speculative interpretations aside, these data demonstrate that impairments in bodily self-monitoring can cause a cross-modal effect on loudness perception, that is additionally modulated with subjective states associated to those impairments.

4.2.3 Explicit and implicit self-other voice discrimination

In the same study, we reported that different degrees of sensorimotor stimulation (synchronous, asynchronous) did not modulate explicit SOVD. Similar was observed in the study of (Salomon et al., 2020), where robotic-sensorimotor stimulation induced errors in auditory-verbal self-monitoring only in schizophrenia patients with passivity experiences. They observed no effect in a healthy control group, but neither in patients without passivity experiences. Thus, only those individuals who already experience misattributions of self-generated actions towards other agents were susceptible to confusing own and other voices under asynchronous sensorimotor stimulation. It might be that self-monitoring impairments induced by the robot are simply not strong enough to in-

duce such an effect in individuals without inherent self-monitoring deficits, such as passivity experiences. Compared to the study of (Salomon et al., 2020), my study involved a much more sensitive task employing voice morphing technology (Kawahara et al., 2013), that should be capable of capturing even small changes in SOVD (in an adequately powered sample).

Interestingly, sensorimotor stimulation modulated identity of voices in the task employed in Study 5 – voice detection. Relating the two studies, the task of Study 5 could be seen as an implicit SOVD task. Participants were asked to report whether they heard a voice in the noise, and not which voice. We can not know whether participants were aware of the identity of the voices they were hearing, especially since the voices were presented at their hearing thresholds, but what is important is that they did not orient their attention towards the identity. It is possible that directly focusing on who is speaking (such as in explicit SOVD task) involves higher-level cognitive processes, which can not be affected by low-level sensorimotor modulations imposed by the robot. Accordingly, lower-level auditory processes, such as loudness perception in Study 3 or voice detection in Study 5, could interact with sensorimotor stimulation more easily.

The fact that identity was modulated in the implicit SOVD task further resembles the so-called self-advantage phenomenon, where task performance that does not explicitly assess self-relatedness is modulated by self-relatedness of the stimuli (Ma & Han, 2010). For example, in the study of (Frassinetti, Maini, Romualdi, Galante, & Avanzi, 2008), when asked to indicate which of two images (high or low) matched a central target stimulus, participants performed better with self rather than other' body-parts. By contrast, a lack of this facilitation was observed when the task employed an explicit judgement about which of two images contains their body part (Frassinetti, Ferri, Maini, Benassi, & Gallese, 2011). These authors have proposed that a sensorimotor body-representation is engaged in the implicit, but not in the explicit recognition of one's own body-parts (Ferri, Frassinetti, Ardizzi, Costantini, & Gallese, 2012) and have extended differences in implicit and explicit recognition to self-voice perception (Candini et al., 2014).

4.2.4 Dependency on breathing

Explicit SOVD was affected by robotic-sensorimotor stimulation when seen from the perspective of interoceptive autonomic processing. Specifically, we observed an increase in SOVD task performance during inspiration, compared to expiration phase of breathing, that was present only during asynchronous sensorimotor stimulation. Breathing has only recently been related to cognitive processing, and only a few studies reported a modulation of behavioral task performance as a function of respiration phase (Heck, Kozma, & Kay, 2019; Nakamura, Fukunaga, & Oku, 2018; Perl et al., 2019; Zelano et al., 2016). Interestingly, all studies reported an increase in employed task performance during inspiration phase. These authors proposed that inspiration-driven advantage arises either from breathing-associated neural entrainment of task-relevant regions (e.g., more activity in hippocampus during inspiration compared to expiration was related to a memory task (Zelano et al., 2016)) or from changes in connectivity patterns in inspiration compared to expiration, which optimize processing of incoming information of all kinds (Perl et al., 2019).

Our results extend this work on breathing-dependent cognition to self-processing, but they also provide interesting insights in the relationship between sensorimotor stimulation and voice perception. Specifically, we observed that inspiration improved SOVD specifically during asynchronous stimulation and more prominently in individuals who experienced somatic passivity. We propose that no effect of breathing during synchronous stimulation might represent a form of a ceiling effect – i.e. as self-monitoring is not 'shifted away' towards external agents, subtle differences in SOVD introduced by breathing could not be registered. During asynchronous stimulation, however, where BSC is altered (or arguably even impaired) and participants experience other-agency sensations, the same effects of breathing might introduce significant and noticeable differences.

4.2.5 Vocal false alarms

Inspired by the selectivity of the loudness effect in Study 3 to quiet voices, I was wondering what would happen if I make the voices really quiet. Building up on studies that associated hallucinations in healthy individuals and increased false alarm rates in auditory detection tasks (Barkus et al., 2011; Moseley, Fernyhough, & Ellison, 2014; Powers et al., 2017), I designed a voice-detection task with self- and other-voice stimuli presented at individual hearing thresholds that is conducted under concomitant sensorimotor stimulation. In two experiments with the same sample size, I observed and replicated a higher rate of vocal false alarms during asynchronous, compared to synchronous stimulation, and additionally an interaction between sensorimotor stimulation and voice identity on false alarms. This interaction revealed that in other-voice blocks there were more false alarms with asynchronous stimulation, whereas in self-voice blocks, false alarms were increased with synchronous stimulation. In addition, the increase in false alarms was positively related to delusion proneness. To the best of my knowledge, this is the first report of hallucinations being induced in a laboratory directly as a consequence of self-monitoring impairments. Namely, all the evidence relating self-monitoring

to hallucinogenesis is correlational (for a review of a certain subset of such studies see (Whitford, 2019)). Results presented in Study 5 are thus of vast clinical relevance, as they shed new light on the mechanisms underlying AVH, providing causal evidence that associates self-monitoring impairments to AVH.

However, what remains unclear from these findings is the interpretation of the observed interaction effect. Namely, before conducting the first study, we expected to observe a higher false alarm rate during asynchronous stimulation, that has been previously associated to hallucinations in healthy individuals (Blanke et al., 2014). Additionally, as asynchronous stimulation has further been related to somatic passivity and presence hallucination, both of which introduce a form of alien agency, we predicted that this increase might be accentuated during other-voice blocks. We did not expect to observe, let alone replicate an effect in the synchronous condition. Since the beginning of the hallucination-oriented work in our lab, synchronous condition has been considered as a control for asynchronous condition, as it contained no temporal conflict. It was purposefully designed to match the cognitive load to the hallucination-related asynchronous stimulation in the closest possible manner – i.e. so that participants perform equivalent movements and perceive equivalent touches, albeit without hallucinating. Thus, in a search for an explanation of the effect observed in synchronous stimulation, I considered reports of inducing hallucinations in healthy participants based on principles other than self-monitoring impairments.

4.2.6   The strong-prior account

There have been other reports of inducing auditory hallucinations in a controlled laboratory environment, moreover in a non-hallucinating population. This was mostly achieved through conditioning paradigms (Pavlov, 1928), where experimentally manipulated expectations of target presence (e.g. a tone) – built upon learned associations between the target and accompanying stimuli (e.g. seeing an illuminated bulb) – drive reports of perceiving the target even in its absence (Barber & Calverley, 1964; Davies, Davies, & Bennett, 1982; Ellson, 1941; Seashore, 1895). Another interesting approach involved transcranial direct current stimulation of left posterior superior temporal gyrus, cortical area associated with AVH (Moseley, Fernyhough, & Ellison, 2014). In these studies, laboratory-induced auditory hallucinations were objectively quantified as the false alarm rate in signal detection tasks (Green & Swets, 1966). It has further been shown that hallucination-prone individuals (Barkus et al., 2011; Barkus, Stirling, Hopkins, McKie, & Lewis, 2007), as well as in non-clinical voice-hearers (Powers et al., 2017) tend to report higher false alarms in auditory detection tasks, consolidating the role of false alarms as a measure of laboratory-induced hallucinations.

Thus, the findings of Study 5 have been discussed also through the prism of malfunctioned prior weighting. It is important to note that our paradigm did not contain Bayesian elements nor it manipulated stimulus probabilities and that what follows are just conceptual speculations. A potential support to the strong-prior account of our data might be the observed relationship between experimentally-induced AVH with an increase in PDI score, a measure of delusional ideation in the general population (Peters, Joseph, Day, & Qarety, 2004). PDI it has been indicative of excessive reliance on priors (Schmack et al., 2013; Teufel et al., 2015) and, more related to our findings, to self-monitoring deficits (Teufel et al., 2010). The question that follows is how could self-monitoring impairments be related to strong priors and what would those priors be?

I believe that these data, and especially the bidirectionality of the observed interaction, could be in concordance with the model proposed by (Leptourgos & Corlett, 2020). This model reconciles self-monitoring and strong-prior accounts by suggesting that (not) explaining-away the sensory input by feedforward predictions affects selective agency-related priors, such as self-attribution priors (Farrer, Franck, Paillard, & Jeannerod, 2003; Hauser et al., 2011; Tsakiris et al., 2005). Such attribution biases might be combined with the 'auditory' prior – a simple expectation to hear the same type of voice following its repetitive presentation in the recent past – into selective false vocal reports. Specifically, an increase in precision of a self-attribution prior (that results from temporal congruence between poking movements and somatosensory feedback) could render other systems more sensitive to self-related stimuli. When exposed to a repetitive presentation of self-voices, the auditory system expects to hear self-voices again in the near future. Expectation to hear self-voice combined with a concomitant increase in sensitivity to self-related stimuli might result in false percepts of self-voice stimuli under situations of uncertainty. The same could apply to other-attribution bias combined with an expectation to hear other voices.

4.3   Conclusion and outlook

To conclude, the empirical work I presented in this thesis serves as a scaffold for a new understanding of self-voice phenomenon by identifying its neural correlates and portraying it as a fundamentally multimodal construct. It introduces new methods that enable a more ecological approach to conducting self-voice research and demonstrate that AVH can be induced in healthy individuals in a controlled laboratory environment. Finally, it is of vast clinical relevance as it sheds new light on the etiology of AVH, providing causal evidence of AVH induction resulting from self-monitoring impairments.

As this thesis exclusively investigated behavioral and neural mechanisms underlying passive self-voice perception and SOVD, future work should extend these paradigms to active self-voice perception, i.e. involving voice production. Voice production opens the door to self-monitoring mechanisms directly related to speaking (Burnett, Freedland, Larson, & Hain, 1998; Houde & Jordan, 1998; Jones & Munhall, 2000), thus it would be interesting to compare active and passive versions of the SOVD task and additionally explore the role of bone conduction. I predict that, due to feedforward predictions about the sensory consequences of our voice during speech production, healthy participants would exhibit a self-bias during the active, compared to the passive SOVD task. Moreover, I expect that this bias would be reduced with concomitant robotic-asynchronous and amplified with concomitant synchronous stimulation, thereby revealing interesting interactions between the two simultaneously active sensorimotor loops – one associated to speaking, and another to bodily somatosensory-motor predictions. As many participants complained about the SOVD task being very difficult, and yet performed well, it would further be interesting to investigate second-order, metacognitive performance related to SOVD.

Considering the relevance of this work to AVH, I envision three possible directions future research endeavors could take. First, the next logical step would be to formalize these findings by fitting this data to an appropriate computational model. New experiments could be specifically designed to assess differences between synchronous and asynchronous stimulations with respect to predefined parameters of such a model. For example, in the study of (Powers et al., 2017), a specific parameter of a Hierarchical Gaussian Model indicated differences in reliance on priors between voice-hearers and non-hallucinators. Second, a relationship between SOVD deficits and alterations in prior weighting should be further investigated. If the two accounts are indeed related, manipulating one should change the other. Thus, future studies could aim at causally inducing SOVD impairments in healthy non-hallucinators by manipulating the precision of incoming sensory evidence relative to the precision of prior knowledge (Marshall et al., 2016). This could be achieved pharmacologically (Linster & Cleland, 2002; Moran et al., 2013) or by imposing strong priors about the voice identity through carefully-designed experimental paradigms (e.g. through learned associations to faces). Finally, as the initial motivation of this work was to help demystifying AVH, these findings should be verified in voice-hearers, in order to pinpoint specific SOVD impairments occurring in this population and relate them to clinical assessments of symptom severity. This would lead to a deeper understanding of AVH, fortifying its relatedness to self-monitoring, thereby paving the way towards the development of novel therapeutic approaches.

References

Adams, R. A., Brown, H. R., & Friston, K. J. (2015). Bayesian inference, predictive coding and delusions. Avant, V(3), 51–88. https://doi.org/10.12849/50302014.0112.0004

Adler, D., Herbelin, B., Similowski, T., & Blanke, O. (2014). Breathing and sense of self: Visuo-respiratory conflicts alter body self-consciousness. Respiratory Physiology & Neurobiology, 203, 68–74. https://doi.org/10.1016/j.resp.2014.08.003

Alderson-Day, B., & Fernyhough, C. (2016). Auditory verbal hallucinations: Social, but how? Journal of Consciousness Studies, 23(7–8), 163–194. https://doi.org/10.1007/978-1-4614-0959-5_9

Alderson-Day, B., Lima, C. F., Evans, S., Krishnan, S., Shanmugalingam, P., Fernyhough, C., & Scott, S. K. (2017). Distinct processing of ambiguous speech in people with non-clinical auditory verbal hallucinations. Brain, 140(9), 2475–2489. https://doi.org/10.1093/brain/awx206

Allen, P., Aleman, A., & McGuire, P. K. (2007). Inner speech models of auditory verbal hallucinations: Evidence from behavioural and neuroimaging studies. International Review of Psychiatry, 19(4), 407–415. https://doi.org/10.1080/09540260701486498

Allen, P. P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M., Johns, L. C., & McGuire, P. K. (2005). Neural correlates of the misattribution of self-generated speech. Human Brain Mapping, 26(1), 44–53. https://doi.org/10.1002/hbm.20120

Allen, P. P., Johns, L. C., Fu, C. H. Y., Broome, M. R., Vythelingum, G. N., & McGuire, P. K. (2004). Misattribution of external speech in patients with hallucinations and delusions. Schizophrenia Research, 69(2–3), 277–287. https://doi.org/10.1016/j.schres.2003.09.008

Alzueta, E., Melcón, M., Poch, C., & Capilla, A. (2019). Is your own face more than a highly familiar face? Biological Psychology, 142(January), 100–107. https://doi.org/10.1016/j.biopsycho.2019.01.018

Ameller, A., Dereux, A., Dubertret, C., Vaiva, G., Thomas, P., & Pins, D. (2015). 'What is more familiar than I? Self, other and familiarity in schizophrenia. Schizophrenia Research, 161(2–3), 501–505. https://doi.org/10.1016/j.schres.2014.12.007

Ameller, A., Picard, A., D'Hondt, F., Vaiva, G., Thomas, P., & Pins, D. (2017). Implicit recognition of familiar and unfamiliar faces in schizophrenia: A study of the skin conductance response in familiarity disorders. Frontiers in Psychiatry, 8(SEP), 1–11. https://doi.org/10.3389/fpsyt.2017.00181

Asai, T. (2016). Self is "other", other is "self"_ poor self-other discriminability explains schizotypal twisted agency judgment. Psychiatry Research, 246(February), 593–600. https://doi.org/10.1016/j.psychres.2016.10.082

Aspell, J. E., Heydrich, L., Marillier, G., Lavanchy, T., Herbelin, B., & Blanke, O. (2013). Turning Body and Self Inside Out: Visualized Heartbeats Alter Bodily Self-Consciousness and Tactile Perception. Psychological Science, 24(12), 2445–2453. https://doi.org/10.1177/0956797613498395

Barber, T. X., & Calverley, D. S. (1964). An experimental study of "hypnotic" (auditory and visual) hallucinations. The Journal of Abnormal and Social Psychology, 68(1), 13.

Barkus, E., Smallman, R., Royle, N., Barkus, C., Lewis, S., & Rushe, T. (2011). Auditory false perceptions are mediated by psychosis risk factors. Cognitive Neuropsychiatry, 16(4), 289–302. https://doi.org/10.1080/13546805.2010.530472

Barkus, E., Stirling, J., Hopkins, R., McKie, S., & Lewis, S. (2007). Cognitive and neural processes in non-clinical auditory hallucinations. British Journal of Psychiatry, 191(S51), 76–81. https://doi.org/10.1192/bjp.191.51.s76

Bauer, S. M., Schanda, H., Karakula, H., Olajossy-Hilkesberger, L., Rudaleviciene, P., Okribelashvili, N., … Stompe, T. (2011). Culture and the prevalence of hallucinations in schizophrenia. Comprehensive Psychiatry, 52(3), 319–325. https://doi.org/10.1016/j.comppsych.2010.06.008

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. Psychological Research, 74(1), 110–120. https://doi.org/10.1007/s00426-008-0185-z

Békésy, G. V. (1949). The Structure of the Middle Ear and the Hearing of One's Own Voice by Bone Conduction. Journal of the Acoustical Society of America, 21(3), 217–232. https://doi.org/10.1121/1.1906501

Bernasconi, F., Blondiaux, E., Potheegadoo, J., Stripeikyte, G., Bejr-kasem, H., Pagonabarraga, J., … Blanke, O. (2020). Sensorimotor hallucinations in Parkinson' s disease. BioRxiv.

Blakemore, S. J., Smith, J., Steel, R., Johnstone, E. C., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. Psychological Medicine, 30(5), 1131–1139. https://doi.org/10.1017/S0033291799002676

Blakemore, S J, Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? Neuroreport, 11(11), R11-6. https://doi.org/10.1586/14737175.7.10.1337

Blakemore, Sarah J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. Journal of Cognitive Neuroscience, 11(5), 551–559. https://doi.org/10.1162/089892999563607

Blakemore, Sarah J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. Nature Neuroscience, 1(7), 635–640. https://doi.org/10.1038/2870

Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. Neuroscience and Biobehavioral Reviews, 47, 717–734. https://doi.org/10.1016/j.neubiorev.2014.10.022

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. Nature Reviews Neuroscience, 13(8), 556–571. https://doi.org/10.1038/nrn3292

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. Trends in Cognitive Sciences, 13(1), 7–13. https://doi.org/10.1016/j.tics.2008.10.003

Blanke, O., Pozeg, P., Hara, M., Heydrich, L., Serino, A., Yamamoto, A., … Rognini, G. (2014). Neurological and robot-controlled induction of an apparition. Current Biology, 24(22), 2681–2686. https://doi.org/10.1016/j.cub.2014.09.049

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness. Neuron, 88(1), 145–166. https://doi.org/10.1016/j.neuron.2015.09.029

Blondiaux, E. (2020). Hallucination Engineering: From Neuroscience Robotics to Hallucinations in Health and Disease. EPFL.

Botvinick, M., & Cohen, J. D. (1998). Rubber hand 'feels' what eyes see. Nature, 391(February), 756.

Braun, N., Debener, S., Spychala, N., Bongartz, E., Sörös, P., Müller, H. H. O., & Philipsen, A. (2018). The senses of agency and ownership: A review. Frontiers in Psychology, 9(535), 1–17. https://doi.org/10.3389/fpsyg.2018.00535

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. The Journal of the Acoustical Society of America, 103(6), 3153–3161. https://doi.org/10.1121/1.423073

Candini, M., Zamagni, E., Nuzzo, A., Ruotolo, F., Iachini, T., & Frassinetti, F. (2014). Who is speaking? Implicit and explicit self and other voice recognition. Brain and Cognition, 92, 112–117. https://doi.org/10.1016/j.bandc.2014.10.001

Canzoneri, E., Herbelin, B., Blanke, O., & Serino, A. (2016). Conceptual processing is referenced to the experienced location of the self, not to the location of the physical body. Cognition, 154, 182–192. https://doi.org/10.1016/j.cognition.2016.05.016

Cassidy, C. M., Balsam, P. D., Weinstein, J. J., Rosengard, R. J., Slifstein, M., Daw, N. D., … Horga, G. (2018). A Perceptual Inference Mechanism for Hallucinations Linked to Striatal Dopamine. Current Biology, 28(4), 503–514. https://doi.org/10.1016/j.cub.2017.12.059

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2016). The effects of stimulus complexity on the preattentive processing of self-generated and nonself voices: An ERP study. Cognitive, Affective and Behavioral Neuroscience, 16(1), 106–123. https://doi.org/10.3758/s13415-015-0376-1

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2018). Stimulus complexity matters when you hear your own voice: Attention effects on self-generated voice processing. International Journal of Psychophysiology, 133(June), 66–78. https://doi.org/10.1016/j.ijpsycho.2018.08.007

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-day, B., Schmack, K., & Iii, A. R. P. (2019). Hallucinations and Strong Priors. Trends in Cognitive Sciences, 23(2), 114–127. https://doi.org/10.1016/j.tics.2018.12.001

Critchley, M. (1955). The idea of a presence. Acta Psychiatrica Scandinavica, 30(1–2), 155–168. https://doi.org/10.1111/j.1600-0447.1955.tb06055.x

Daalman, K., Diederen, K. M. J., Hoekema, L., Van Lutterveld, R., & Sommer, I. E. C. (2016). Five year follow-up of non-psychotic adults with frequent auditory verbal hallucinations: Are they still healthy? Psychological Medicine, 46, 1897–1907. https://doi.org/10.1017/S0033291716000386

David, N., Newen, A., & Vogeley, K. (2008). The "sense of agency" and its underlying cognitive and neural mechanisms. Consciousness and Cognition, 17(2), 523–534. https://doi.org/10.1016/j.concog.2008.03.004

Davies, P., Davies, G. L., & Bennett, S. (1982). An effective paradigm for conditioning visual perception in human subjects. Perception, 11(6), 663–669.

Dobrev, I., Sim, J. H., Stenfelt, S., Ihrle, S., Gerig, R., Pfiffner, F., … Röösli, C. (2017). Sound wave propagation on the human skull surface with bone conduction stimulation. Hearing Research, 355, 1–13. https://doi.org/10.1016/j.heares.2017.07.005

Douglas, W., & Gibbins, K. (1983). Inadequacy of voice recognition as a demonstration of self-deception. Journal of Personality and Social Psychology, 44(3), 589–592. https://doi.org/10.1037/0022-3514.44.3.589

Ellson, D. G. (1941). Hallucinations produced by sensory conditioning. Journal of Experimental Psychology 28, No. 1 (1941): 1., 28(1), 1–20.

Emami, S. F., Pourbakht, A., Sheykholeslami, K., Kamali, M., Behnoud, F., & Daneshi, A. (2012). Vestibular Hearing and Speech Processing. ISRN Otolaryngology, 2012, 1–7. https://doi.org/10.5402/2012/850629

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. European Journal of Neuroscience, 15, 399–402. https://doi.org/10.1046/j.0953-816x.2001.01874.x

Farrer, C., Bouchereau, M., Jeannerod, M., & Franck, N. (2008). Effect of Distorted Visual Feedback on the Sense of Agency. Behavioural Neurology, 19(1–2), 53–57. https://doi.org/10.1155/2008/425267

Farrer, C., Franck, N., Paillard, J., & Jeannerod, M. (2003). The role of proprioception in action recognition. Consciousness and Cognition, 12(4), 609–619. https://doi.org/10.1016/S1053-8100(03)00047-3

Feinberg, I. (1978). Efference copy and corollary discharge: implications for thinking and its disorders. Schizophrenia Bulletin, 4(4), 636–640. https://doi.org/10.1093/schbul/4.4.636

Ferri, F., Frassinetti, F., Ardizzi, M., Costantini, M., & Gallese, V. (2012). A sensorimotor network for the bodily self. Journal of Cognitive Neuroscience, 24(7), 1584–1595. https://doi.org/10.1162/jocn_a_00230

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nature Reviews Neuroscience, 10(1), 48–58. https://doi.org/10.1038/nrn2536

Ford, J. M., Gray, M., Faustman, W. O., Roach, B. J., & Mathalon, D. H. (2007). Dissecting corollary discharge dysfunction in schizophrenia. Psychophysiology, 44(4), 522–529. https://doi.org/10.1111/j.1469-8986.2007.00533.x

Ford, J. M., & Mathalon, D. H. (2005). Corollary discharge dysfunction in schizophrenia: Can it explain auditory hallucinations? International Journal of Psychophysiology, 58(2-3 SPEC. ISS.), 179–189. https://doi.org/10.1016/j.ijpsycho.2005.01.014

Ford, J. M., Mathalon, D. H., Heinks, T., Kalba, S., Faustman, W. O., & Roth, W. T. (2001). Neurophysiological Evidence of Corollary Discharge Dysfunction in Schizophrenia. Am J Psychiatry, 158(12), 2069–2071. Retrieved from http://ajp.psychiatryonline.org/cgi/reprint/158/12/2069

Ford, J. M., Mathalon, D. H., Whitfield, S., Faustman, W. O., & Roth, W. T. (2002). Reduced communication between frontal and temporal lobes during talking in schizophrenia. Biological Psychiatry, 51(6), 485–492. https://doi.org/10.1016/S0006-3223(01)01335-X

Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2007). Synch before you speak: Auditory hallucinations in schizophrenia. American Journal of Psychiatry, 164(3), 458–466. https://doi.org/10.1176/appi.ajp.164.3.458

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., … Murray, M. M. (2001). Auditory-Somatosensory Multisensory Processing in Auditory Association Cortex: An fMRI Study. Journal of Neurophysiology, 88, 540–543. https://doi.org/10.1152/jn.00694.2001

Frassinetti, F., Ferri, F., Maini, M., Benassi, M. G., & Gallese, V. (2011). Bodily self: An implicit knowledge of what is explicitly unknown. Experimental Brain Research, 212(1), 153–160. https://doi.org/10.1007/s00221-011-2708-x

Frassinetti, F., Maini, M., Romualdi, S., Galante, E., & Avanzi, S. (2008). Is it Mine? Hemispheric Asymmetries in Corporeal Self-recognition. Journal of Cognitive Neuroscience, 20(8), 1507–1516. https://doi.org/10.1162/jocn.2008.20067

Frith, C. (1992). The cognitive neuropsychology of schizophrenia. Psychology press.

Frith, C. D. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. Psychological Medicine, 17(3), 631–648. https://doi.org/10.1017/S0033291700025873

Frith, C. D. (1992). The cognitive neuropsychology of schizophrenia. Erlbaum. https://doi.org/10.1016/0028-3932(93)90120-O

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 355(1404), 1771–1788. https://doi.org/10.1098/rstb.2000.0734

Frith, C. D., & Done, D. J. (1988). Towards a neuropsychology of schizophrenia. British Journal of Psychiatry, 153(4), 437–443. https://doi.org/10.1192/bjp.153.4.437

Frith, Chris D., Blakemore, S. J., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. Brain Research Reviews, 31(2–3), 357–363. https://doi.org/10.1016/S0165-0173(99)00052-1

Frith, Christopher D., & Done, D. J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. Psychological Medicine, 19(2), 359-363. https://doi.org/10.4324/9781315630502

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. Trends in Cognitive Sciences, 4(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. Psychological Bulletin, 131(1), 76–97. https://doi.org/10.1037/0033-2909.131.1.76

Gould, L. N. (1948). Verbal hallucinations and activity of vocal musculature; an electromyographic study. The American Journal of Psychiatry, 105, 367–72. https://doi.org/10.1176/ajp.105.5.367

Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., & Bruneau, N. (2015). Is my voice just a familiar voice? An electrophysiological study. Social Cognitive and Affective Neuroscience, 10(1), 101–105. https://doi.org/10.1093/scan/nsu031

Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., Camus, V., & Bruneau, N. (2013). My voice or yours? An electrophysiological study. Brain Topography, 26(1), 72–82. https://doi.org/10.1007/s10548-012-0233-2

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. (J. A. Swets, Ed.), New York: Wiley. https://doi.org/10.1901/jeab.1969.12-475

Green, M. F., & Kinsbourne, M. (1990). Subvocal activity and auditory hallucinations: Clues for behavioral treatments? Schizophrenia Bulletin, 16(4), 617–625.

Green, P., & Preston, M. (1981). Reinforcement of vocal correlates of auditory hallucinations by auditory feedback: a case study. The British Journal of Psychiatry, 139(3), 204–208.

Greenlee, J. D. W., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., … Howard, M. A. (2011). Human auditory cortical activation during self-vocalization. PLoS ONE, 6(3). https://doi.org/10.1371/journal.pone.0014744

Gregory, R. L. (1973). The confounded eye. Illusion in Nature and Art, 49–96. Retrieved from https://ci.nii.ac.jp/naid/10022399161/

Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. Journal of Personality and Social Psychology, 37(2), 147–169. https://doi.org/10.1037/0022-3514.37.2.147

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. Nature Neuroscience, 5(4), 382–385. https://doi.org/10.1038/nn827

Hanley, J. R., Smith, S. T., & Hadfield, J. (1998). I Recognise You but I Can't Place You: An Investigation of Familiar-only Experiences during Tests of Voice and Face Recognition. Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 51(1), 179–195. https://doi.org/10.1080/713755751

Hara, M., Rognini, G., Evans, N., Blanke, O., Yamamoto, A., Bleuler, H., & Higuchi, T. (2011). A novel approach to the manipulation of body-parts ownership using a bilateral master-slave system. In IEEE International Conference on Intelligent Robots and Systems (pp. 4664–4669). https://doi.org/10.1109/IROS.2011.6048519

Harkavy-Friedman, J. M., Kimhy, D., Nelson, E. A., Venarde, D. F., Malaspina, D., & Mann, J. J. (2003). Suicide attempts in schizophrenia: the role of command auditory hallucinations for suicide. The Journal of Clinical Psychiatry., 64(8), 871–874.

Hauser, M., Knoblich, G., Repp, B. H., Lautenschlager, M., Gallinat, J., Heinz, A., & Voss, M. (2011). Altered sense of agency in schizophrenia and the putative psychotic prodrome. Psychiatry Research, 186(2–3), 170–176. https://doi.org/10.1016/j.psychres.2010.08.003

Heck, D. H., Kozma, R., & Kay, L. M. (2019). The rhythm of memory: how breathing shapes memory function. Journal of Neurophysiology, 122(2), 563–571. https://doi.org/10.1152/jn.00200.2019

Helmholtz, H. von. (1866). Treatise on Physiological Optics. Book.

Heydrich, L., Aspell, J. E., Marillier, G., Lavanchy, T., Herbelin, B., & Blanke, O. (2018). Cardio-visual full body illusion alters bodily self-consciousness and tactile processing in somatosensory cortex. Scientific Reports, 8(1), 1–8. https://doi.org/10.1038/s41598-018-27698-2

Hoffman, R. E., Fernandez, T., Pittman, B., & Hampson, M. (2011). Elevated functional connectivity along a corticostriatal loop and the mechanism of auditory/verbal hallucinations in patients with schizophrenia. Biological Psychiatry, 69(5), 407–414.

Holzman, P. S., Rousey, C., & Snyder, C. (1966). On listening to one's own voice: Effects on psychophysiological responses and free associations. Journal of Personality and Social Psychology, 4(4), 432–441. https://doi.org/10.1037/h0023790

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. Science, 279(5354), 1213–1216. https://doi.org/10.1126/science.279.5354.1213

Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. Psychological Bulletin, 139(1), 133–151. https://doi.org/10.1037/a0028566

Hughes, S. M., & Nicholson, S. E. (2010). The processing of auditory and visual recognition of self-stimuli. Consciousness and Cognition, 19(4), 1124–1134. https://doi.org/10.1016/j.concog.2010.03.001

Hurlburt, R. T., Heavey, C. L., & Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. Consciousness and Cognition, 22(4), 1477–1494. https://doi.org/10.1016/j.concog.2013.10.003

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. Proceedings of the National Academy of Sciences, 106(4), 1245–1248. https://doi.org/10.1073/pnas.0810063106

Jacobson, E. (1931). Electrical measurements of neuromuscular states during mental activities: V. Variation of specific muscles contracting during imagination. American Journal of Physiology-Legacy Content, 96(1), 115–121. https://doi.org/10.1152/ajplegacy.1931.96.1.115

Jaspers, K. (1990). Über leibhaftige Bewußtheiten (Bewußtheitstäuschungen), ein psychopathologisches Elementarsymptom. In Gesammelte Schriften zur Psychopathologie (pp. 413–420). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-62027-0_8

Jeannerod, M. (2003). The mechanism of self-recognition in humans. Behavioural Brain Research, 142, 1–15. https://doi.org/10.1016/S0166-4328(02)00384-4

Jeannerod, M. (2006). The origin of voluntary action. History of a physiological concept. Comptes Rendus Biologies, 329(5–6), 354–362. https://doi.org/10.1016/j.crvi.2006.03.017

Jeannerod, M., & Pacherie, E. (2004). Agency, Simulation and Self-identification. Mind and Language, 19(2), 113–146. https://doi.org/10.1111/j.1468-0017.2004.00251.x

Johns, L. C., Cannon, M., Singleton, N., Murray, R. M., Farrell, M., Brugha, T., … Meltzer, H. (2004). Prevalence and correlates of self-reported psychotic symptoms in the British population. The British Journal of Psychiatry, 185(4), 298–305.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. The Journal of the Acoustical Society of America, 108(3), 1246. https://doi.org/10.1121/1.1288414

Kannape, O. A., & Blanke, O. (2012). Agency, gait and self-consciousness. International Journal of Psychophysiology, 83(2), 191–199. https://doi.org/10.1016/j.ijpsycho.2011.12.006

Kannape, O. A., & Blanke, O. (2013). Self in motion: sensorimotor and cognitive mechanisms in gait agency. Journal of Neurophysiology, 110(8), 1837–1847. https://doi.org/10.1152/jn.01042.2012

Kannape, Oliver A., Smith, E. J. T., Moseley, P., Roy, M. P., & Lenggenhager, B. (2019). Experimentally induced limb-disownership in mixed reality. Neuropsychologia, 124(August 2018), 161–170. https://doi.org/10.1016/j.neuropsychologia.2018.12.014

Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q., & Iacoboni, M. (2008). The self across the senses: an fMRI study of self-face and self-voice recognition. Social Cognitive and Affective Neuroscience, 3(3), 218–223. https://doi.org/10.1093/scan/nsn014

Kawahara, H., Morise, M., Banno, H., & Skuk, V. G. (2013). Temporally variable multi-aspect N-way morphing based on interference-free speech representations. In 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2013 (pp. 1–10). IEEE. https://doi.org/10.1109/APSIPA.2013.6694355

Kelleher, I., Lynch, F., Harley, M., Molloy, C., Roddy, S., Fitzpatrick, C., & Cannon, M. (2012). Psychotic symptoms in adolescence index risk for suicidal behavior: findings from 2 population-based case-control clinical interview studies. Archives of General Psychiatry, 69(12), 1277–1283.

Kilteni, K., & Ehrsson, H. H. (2017). Sensorimotor predictions and tool use: Hand-held tools attenuate self-touch. Cognition, 165, 1–9. https://doi.org/10.1016/j.cognition.2017.04.005

Kimura, M., & Yotsumoto, Y. (2018). Auditory traits of "own voice." PLoS ONE, 13(6), 1–16. https://doi.org/10.1371/journal.pone.0199443

Kot, T., & Serper, M. (2002). Increased susceptibility to auditory conditioning in hallucinating schizophrenic patients: A preliminary investigation. Journal of Nervous and Mental Disease, 190(5), 282–288. https://doi.org/10.1097/00005053-200205000-00002

Lehman, A. F., Lieberman, J. A., Dixon, L. B., McGlashan, T. H., Miller, A. L., Perkins, D. O., … Cook, I. (2004). Practice guideline for the treatment of partients with schizophrenia. American Journal of Psychiatry, 161(2 SUPPL), 1–56.

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video Ergo Sum: Manipulating Bodily Self-Consciousness. Science, 317(5841), 1096–1099. https://doi.org/10.1126/science.1143439

Leptourgos, P., & Corlett, P. R. (2020). Embodied Predictions, Agency, and Psychosis. Frontiers in Big Data, 3(August), 27. https://doi.org/10.3389/FDATA.2020.00027

Linster, C., & Cleland, T. A. (2002). Cholinergic modulation of sensory representations in the olfactory bulb. Neural Networks, 15(4–6), 709–717.

Liu, L., Li, W., Li, J., Lou, L., & Chen, J. (2019). Temporal features of psychological and physical self-representation: An ERP study. Frontiers in Psychology, 10(APR). https://doi.org/10.3389/fpsyg.2019.00785

Livesay, J., Liebke, A., Samaras, M., & Stanley, A. (1996). Covert speech behavior during a silent language recitation task. Perceptual and Motor Skills, 83(3_suppl), 1355–1362.

Lœvenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., Dohen, M., Perrier, P., … Perrone-Bertolotti, M. (2018). A cognitive neuroscience view of inner language: to predict and to hear, see, feel. In P. Langland-Hassan & A. Vicente (Eds.), Inner speech: New voices (pp. 131–167). Oxford University Pres.

Ma, Y., & Han, S. (2010). Why we respond faster to the self than to others? An implicit positive association theory of self-advantage during implicit face recognition. Journal of Experimental Psychology: Human Perception and Performance, 36(3), 619. https://doi.org/10.1037/a0015797

MacDonald, P. A., & Paus, T. (2003). The role of parietal cortex in awareness of self-generated movements: A transcranial magnetic stimulation study. Cerebral Cortex, 13(9), 962–967. https://doi.org/10.1093/cercor/13.9.962

Manning, C., Mermagen, T., & Scharine, A. (2017). The effect of sensorineural hearing loss and tinnitus on speech recognition over air and bone conduction military communications headsets. Hearing Research, 349, 67–75. https://doi.org/10.1016/j.heares.2016.10.019

Marshall, L., Mathys, C., Ruge, D., De Berker, A. O., Dayan, P., Stephan, K. E., & Bestmann, S. (2016). Pharmacological fingerprints of contextual uncertainty. PLoS Biology, 14(11), e1002575.

Martikainen, M. H., Kaneko, K. I., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. Cerebral Cortex, 15(3), 299–302. https://doi.org/10.1093/cercor/bhh131

Maurer, D., & Landis, T. (1990). Role of bone conduction in the self-perception of speech. Folia Phoniatrica et Logopaedica, 42(5), 226–229. https://doi.org/10.1159/000266070

McCarthy-Jones, S., Trauer, T., MacKinnon, A., Sims, E., Thomas, N., & Copolov, D. L. (2014). A new phenomenological survey of auditory hallucinations: Evidence for subtypes and implications for theory and practice. Schizophrenia Bulletin, 40(1), 225–235. https://doi.org/10.1093/schbul/sbs156

McGuigan, F. J. (1966). Covert oral behavior and auditory hallucinations. Psychophysiology, 3, 73–80. https://doi.org/10.1111/j.1469-8986.1966.tb02682.x

McGuigan, F. J., & Dollins, A. B. (1989). Patterns of covert speech behavior and phonetic coding. The Pavlovian Journal of Biological Science, 24, 19–26. https://doi.org/10.1007/BF02964527

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264(5588), 746–748.

Mechelli, A., Allen, P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M. J., … McGuire, P. K. (2007). Misattribution of speech and impaired connectivity in patients with auditory verbal hallucinations. Human Brain Mapping, 28(11), 1213–1222. https://doi.org/10.1002/hbm.20341

Menzer, F., Brooks, A., Halje, P., Faller, C., Vetterli, M., & Blanke, O. (2010). Feeling in control of your footsteps: Conscious gait monitoring and the auditory consequences of footsteps. Cognitive Neuroscience, 1(3), 184–192. https://doi.org/10.1080/17588921003743581

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. Neural Networks, 9(8), 1265–1279.

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., & Friston, K. J. (2013). Free energy, precision and learning: the role of cholinergic neuromodulation. Journal of Neuroscience, 33(19), 8227–8236.

Moseley, P., Fernyhough, C., & Ellison, A. (2013). Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. Neuroscience and Biobehavioral Reviews, 37(10), 2794–2805. https://doi.org/10.1016/j.neubiorev.2013.10.001

Moseley, P., Fernyhough, C., & Ellison, A. (2014). The role of the superior temporal lobe in auditory false perceptions: A transcranial direct current stimulation study. Neuropsychologia, 62(1), 202–208. https://doi.org/10.1016/j.neuropsychologia.2014.07.032

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., … Foxe, J. J. (2005). Grabbing your ear: Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cerebral Cortex, 15(7), 963–974. https://doi.org/10.1093/cercor/bhh197

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., … Kojima, S. (2001). Neural substrates for recognition of familiar voices: A PET study. Neuropsychologia, 39(10), 1047–1054. https://doi.org/http://dx.doi.org/10.1016/S0028-3932%2801%2900037-9

Nakamura, N. H., Fukunaga, M., & Oku, Y. (2018). Respiratory modulation of cognitive performance during the retrieval process. PLoS ONE, 13(9), e0204021.

Nayani, T. H., & David, A. S. (1996). The auditory hallucination: A phenomenological survey. Psychological Medicine, 26(1), 177–189. https://doi.org/10.1017/s003329170003381x

Nierula, B., Martini, M., Matamala-Gomez, M., Slater, M., & Sanchez-Vives, M. V. (2017). Seeing an Embodied Virtual Hand is Analgesic Contingent on Colocation. The Journal of Pain. https://doi.org/10.1016/j.jpain.2017.01.003

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain- A meta-analysis of imaging studies on the self. NeuroImage, 31(1), 440–457. https://doi.org/10.1016/j.neuroimage.2005.12.002

Northoff, G., & Qin, P. (2011). How can the brain's resting state activity generate hallucinations? A "resting state hypothesis" of auditory verbal hallucinations. Schizophrenia Research, 127(1–3), 202–214. https://doi.org/10.1016/j.schres.2010.11.009

Olivos, G. (1967). Response delay, psychophysiologic activation, and recognition of one's own voice. Psychosomatic Medicine, 29(5), 433–440. https://doi.org/10.1097/00006842-196709000-00003

Park, H. D., & Blanke, O. (2019). Coupling Inner and Outer Body for Self-Consciousness. Trends in Cognitive Sciences, 23(5), 377–388. https://doi.org/10.1016/j.tics.2019.02.002

Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., & Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. European Journal of Neuroscience. https://doi.org/10.1111/j.1460-9568.1996.tb01187.x

Pavlov, I. P. (1928). Natural science and the brain. In Lectures on Conditioned Reflexes: Twenty-five Years of Objective Study of the Higher Nervous Activity (Behaviour) of Animals (pp. 120–130). Liverwright Publishing Corporation.

Perl, O., Ravia, A., Rubinson, M., Eisen, A., Soroka, T., Mor, N., … Sobel, N. (2019). Human non-olfactory cognition phase-locked with inhalation. Nature Human Behaviour, 3(5), 501–512. https://doi.org/10.1038/s41562-019-0556-z

Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciu, M., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. Behavioural Brain Research, 261, 220–239. https://doi.org/10.1016/j.bbr.2013.12.034

Peters, E., Joseph, S., Day, S., & Qarety, P. (2004). Measuring delusional ideation: the 21-item Peters et al. Delusions Inventory (PDI). Schizophrenia Bulletin, 30(4), 1005–1022.

Pfeiffer, C., Schmutz, V., & Blanke, O. (2014). Visuospatial viewpoint manipulation during full-body illusion modulates subjective first-person perspective. Experimental Brain Research, 232(12), 4021–4033. https://doi.org/10.1007/s00221-014-4080-0

Pompili, M., Amador, X. F., Girardi, P., Harkavy-Friedman, J., Harrow, M., Kaplan, K., … Montross, L. P. (2007). Suicide risk in schizophrenia: learning from the past to change the future. Annals of General Psychiatry, 6(1), 1–22.

Pörschmann, C. (2000). Influences of bone conduction and air conduction on the sound of one's own voice. Acustica, 86(6), 1038–1045.

Powers, A R, Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. Science, 357(August), 596–600. https://doi.org/10.1126/science.aan3458

Powers, Albert R., Kelley, M. S., & Corlett, P. R. (2017). Varieties of voice-hearing: Psychics and the psychosis continuum. Schizophrenia Bulletin, 43(1), 84–98. https://doi.org/10.1093/schbul/sbw133

Pozeg, P., Palluel, E., Ronchi, R., Solcà, M., Al-Khodairy, A. W., Jordan, X., … Blanke, O. (2017). Virtual reality improves embodiment and neuropathic pain caused by spinal cord injury. Neurology, 89(18), 1894–1903. https://doi.org/10.1212/WNL.0000000000004585

Preston, C., Kuper-Smith, B. J., & Ehrsson, H. H. (2016). Owning the body in the mirror: The effect of visual perspective and mirror view on the full-body illusion. Scientific Reports, 5(1), 18345. https://doi.org/10.1038/srep18345

Qin, P., Wang, M., & Northoff, G. (2020). Linking bodily, environmental and mental states in the self — A three-level model based on a meta-analysis. Neuroscience and Biobehavioral Reviews, 115, 77–95. https://doi.org/10.1016/j.neubiorev.2020.05.004

Reinfeldt, S., Östli, P., Håkansson, B., & Stenfelt, S. (2010). Hearing one's own voice during phoneme vocalization—Transmission by air and bone conduction. The Journal of the Acoustical Society of America, 128(2), 751–762. https://doi.org/10.1121/1.3458855

Ronchi, R., Heydrich, L., Serino, A., & Blanke, O. (2017). Illusory hand ownership in a patient with personal neglect for the upper limb, but no somatoparaphenia. Journal of Neuropsychology, 1–21. https://doi.org/10.1111/jnp.12123

Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P., & Belin, P. (2008). Investigations of hemispheric specialization of self-voice recognition. Brain and Cognition, 68(2), 204–214. https://doi.org/10.1016/j.bandc.2008.04.007

Sackeim, H. A., & Gur, R. C. (1978). Self-Deception, Self-Confrontation, and Consciousness. In Consciousness and Self-Regulation (pp. 139–197). Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-2571-0_4

Salomon, R., Noel, J., Łukowska, M., Faivre, N., Metzinger, T., Serino, A., & Blanke, O. (2017). Unconscious integration of multisensory bodily inputs in the peripersonal space shapes bodily self-consciousness. Cognition, 166, 174–183. https://doi.org/10.1016/j.cognition.2017.05.028

Salomon, R., Progin, P., Griffa, A., Rognini, G., Do, K. Q., Conus, P., … Blanke, O. (2020). Sensorimotor Induction of Auditory Misattribution in Early Psychosis. Schizophrenia Bulletin. https://doi.org/10.1093/schbul/sbz136

Sartorius, N., Jablensky, A., Korten, A., Ernberg, G., Anker, M., Cooper, J. E., & Day, R. (1986). Early manifestations and first-contact incidence of schizophrenia in different cultures: A preliminary report on the initial evaluation phase of the WHO Collaborative Study on Determinants of Outcome of Severe Mental Disorders. Psychological Medicine, 16(4), 909-928. https://doi.org/10.1017/S0033291700011910

Sass, Louis A, & Parnas, J. (2003). Schizophrenia, consciousness, and the self. Schizophrenia Bulletin, 29(3), 427–444.

Sass, Louis Arnorsson, & Parnas, J. (2001). Phenomenology of Self-Disturbances in Schizophrenia: Some Research Findings and Directions. Philosophy, Psychiatry, & Psychology, 8(4), 347–356. https://doi.org/10.1353/ppp.2002.0027

Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. Consciousness and Cognition, 17(4), 1219–1227. https://doi.org/10.1016/j.concog.2008.01.003

Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: Discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. Cognition, 94(3), 241–255. https://doi.org/10.1016/j.cognition.2004.04.003

Scandola, M., Aglioti, S. M., Pozeg, P., Avesani, R., & Moro, V. (2017). Motor imagery in spinal cord injured people is modulated by somatotopic coding, perspective taking, and post-lesional chronic pain. Journal of Neuropsychology. https://doi.org/10.1111/jnp.12098

Schafer, E. W. P., & Marcus, M. M. (1973). Self-Stimulation Alters Human Sensory Brain Responses. Science, 181(4095), 175–177.

Schmack, K., de Castro, A. G. C., Rothkirch, M., Sekutowicz, M., Rössler, H., Haynes, J. D., … Sterzer, P. (2013). Delusions and the role of beliefs in perceptual inference. Journal of Neuroscience, 33(34), 13701–13712. https://doi.org/10.1523/JNEUROSCI.1778-13.2013

Schuerman, W. L., Meyer, A., & McQueen, J. M. (2015). Do we perceive others better than ourselves? A perceptual benefit for noise-vocoded speech produced by an average speaker. PLoS ONE, 10(7), 1–18. https://doi.org/10.1371/journal.pone.0129731

Seashore, C. E. (1895). Measurements of Illusions and Hallucinations in normal life. Reprinted from Studies from the Yale Psychological Laboratory, Vol. 3.

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error Correction, Sensory Prediction, and Adaptation in Motor Control. Annual Review of Neuroscience, 33(1), 89–108. https://doi.org/10.1146/annurev-neuro-060909-153135

Shergill, S. S., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2003). Two Eyes for an Eye: The Neuroscience of Force Escalation. Science, 301(5630), 187–187. https://doi.org/10.1126/science.1085327

Shergill, S. S., Samson, G., Bays, P. M., Frith, C. D., & Wolpert, D. M. (2005). Evidence for sensory prediction deficits in schizophrenia. American Journal of Psychiatry, 162(12), 2384–2386. https://doi.org/10.1176/appi.ajp.162.12.2384

Shuster, L. I. (1998). The perception of correctly and incorrectly produced /r/. Journal of Speech, Language, and Hearing Research, 41(4), 941–950.

Shuster, L. I., & Durrant, J. D. (2003). Toward a better understanding of the perception of self-produced speech. Journal of Communication Disorders, 36(1), 1–11. https://doi.org/10.1016/S0021-9924(02)00132-6

Siemerkus, J., Tomiello, S., & Stephan, K. E. (2019). Bayesian inference and hallucinations in schizophrenia. Brain, 142(8), 2178–2181. https://doi.org/10.1093/brain/awz211

Solcà, M., Ronchi, R., Bello-Ruiz, J., Schmidlin, T., Herbelin, B., Luthi, F., … Blanke, O. (2018). Heartbeat-enhanced immersive virtual reality to treat complex regional pain syndrome. Neurology, 91(5), e1–e11. https://doi.org/10.1212/WNL.0000000000005905

Sommer, I. E., Daalman, K., Rietkerk, T., Diederen, K. M., Bakker, S., Wijkstra, J., & Boks, M. P. M. (2010). Healthy individuals with auditory verbal hallucinations; Who are they? psychiatric assessments of a selected sample of 103 subjects. Schizophrenia Bulletin, 36(3), 633–641. https://doi.org/10.1093/schbul/sbn130

Stenfelt, S. (2011). Acoustic and physiologic aspects of bone conduction hearing. Implantable Bone Conduction Hearing Aids, (71), 10–21. https://doi.org/10.1159/000323574

Stenfelt, S. (2016). Model predictions for bone conduction perception in the human. Hearing Research, 340, 135–143. https://doi.org/10.1016/j.heares.2015.10.014

Stenfelt, S., & Goode, R. L. (2005). Bone-conducted sound: Physiological and clinical aspects. Otology and Neurotology, 26(6), 1245–1261. https://doi.org/10.1097/01.mao.0000187236.10842.d5

Stenner, M. P., Bauer, M., Sidarus, N., Heinze, H. J., Haggard, P., & Dolan, R. J. (2014). Subliminal action priming modulates the perceived intensity of sensory action consequences. Cognition, 130(2), 227–235. https://doi.org/10.1016/j.cognition.2013.11.008

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., … Corlett, P. R. (2018). The Predictive Coding Account of Psychosis. Biological Psychiatry, 84(9), 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015

Stetson, C., Cui, X., Montague, P. R., & Eagleman, D. M. (2006). Motor-Sensory Recalibration Leads to an Illusory Reversal of Action and Sensation. Neuron, 51(5), 651–659. https://doi.org/10.1016/j.neuron.2006.08.006

Sugiura, M., Kawashima, R., Nakamura, K., Okada, K., Kato, T., Nakamura, A., … Fukuda, H. (2000). Passive and active recognition of one's own face. NeuroImage, 11(1), 36–48. https://doi.org/10.1006/nimg.1999.0519

Sui, J., & Humphreys, G. W. (2017). The ubiquitous self: What the properties of self-bias tell us about the self. Annals of the New York Academy of Sciences, 1396(1), 222–235. https://doi.org/10.1111/nyas.13197

Suzuki, K., Garfinkel, S. N., Critchley, H. D., & Seth, A. K. (2013). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. Neuropsychologia, 51(13), 2909–2917. https://doi.org/10.1016/j.neuropsychologia.2013.08.014

Swiney, L., & Sousa, P. (2014). A new comparator account of auditory verbal hallucinations: how motor prediction can plausibly contribute to the sense of agency for inner speech. Frontiers in Human Neuroscience, 8, 675. https://doi.org/10.3389/fnhum.2014.00675

Synofzik, M., & Vosgerau, G. (2012). Beyond the comparator model. Consciousness and Cognition, 21(1), 1–3. https://doi.org/10.1016/j.concog.2012.01.007

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. Consciousness and Cognition, 17(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

Teufel, C., Kingdon, A., Ingram, J. N., Wolpert, D. M., & Fletcher, P. C. (2010). Deficits in sensory prediction are related to delusional ideation in healthy individuals. Neuropsychologia, 48(14), 4169–4172. https://doi.org/10.1016/j.neuropsychologia.2010.10.024

Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., … Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proceedings of the National Academy of Sciences of the United States of America, 112(43), 13401–13406. https://doi.org/10.1073/pnas.1503916112

Todd, N. P. M. A., Cody, F. W. J., & Banks, J. R. (2000). A saccular origin of frequency tuning in myogenic vestibular evoked potentials?: implications for human responses to loud sounds. Hearing Research, 141(1–2), 180–188. https://doi.org/10.1016/S0378-5955(99)00222-1

Tonndorf, J. (1976). Bone conduction. In Auditory System (pp. 37–84). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-66082-5_13

Tranel, D., & Damasio, A. R. (1985). Knowledge without awareness: An autonomic index of facial recognition by prosopagnosics. Science, 228(4706), 1453–1454.

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. Nature, 423(6942), 866–869. https://doi.org/10.1038/nature01710

Tsakiris, M. (2017). The multisensory basis of the self: From body to identity to others. The Quarterly Journal of Experimental Psychology, 70(4), 597–609. https://doi.org/10.1080/17470218.2016.1181768

Tsakiris, M., & Haggard, P. (2005). Experimenting with the acting self. Cognitive Neuropsychology, 22(3–4), 387–407. https://doi.org/10.1080/02643290442000158

Tsakiris, M., Haggard, P., Franck, N., Mainy, N., & Sirigu, A. (2005). A specific role for efferent information in self-recognition. Cognition, 96(3), 215–231. https://doi.org/10.1016/j.cognition.2004.08.002

Uddin, L. Q., Kaplan, J. T., Molnar-Szakacs, I., Zaidel, E., & Iacoboni, M. (2005). Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: An event-related fMRI study. NeuroImage, 25(3), 926–935. https://doi.org/10.1016/j.neuroimage.2004.12.018

Vurma, A. (2014). The timbre of the voice as perceived by the singer him-/herself. Logopedics Phoniatrics Vocology, 39(1), 1–10. https://doi.org/10.3109/14015439.2013.775334

Weiskrantz, L., Elliott, J., & Darlington, C. (1971). Preliminary observations on tickling oneself. Nature, 230(5296), 598–599. https://doi.org/10.1038/230598a0

Weiss, C., Herwig, A., & Schütz-Bosbach, S. (2011a). The self in action effects: Selective attenuation of self-generated sounds. Cognition, 121(2), 207–218. https://doi.org/10.1016/j.cognition.2011.06.011

Weiss, C., Herwig, A., & Schütz-Bosbach, S. (2011b). The self in social interactions: Sensory attenuation of auditory action effects is stronger in interactions with others. PLoS ONE, 6(7), 16–18. https://doi.org/10.1371/journal.pone.0022723

Wheatstone, C. (1827). Experiments on audition. Quarterly Journal of Science, Literature and Art, 24, 67–72.

Whitford, T. J. (2019). Speaking-Induced Suppression of the Auditory Cortex in Humans and Its Relevance to Schizophrenia. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 4(9), 791–804. https://doi.org/10.1016/j.bpsc.2019.05.011

Whitford, T. J., Mitchell, A. M., & Mannion, D. J. (2017). The ability to tickle oneself is associated with level of psychometric schizotypy in non-clinical individuals. Consciousness and Cognition, 52(May), 93–103. https://doi.org/10.1016/j.concog.2017.04.017

Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., & Grodd, W. (1996). Functional lateralization of speech production at primary motor cortex: A fMRI study. NeuroReport, 7, 2791–6. https://doi.org/10.1097/00001756-199611040-00077

Wilkinson, S. (2014). Accounting for the phenomenology and varieties of auditory verbal hallucination within a predictive processing framework. Consciousness and Cognition, 30, 142–155. https://doi.org/10.1016/j.concog.2014.09.002

Wolpert, D. M. (1997). Computational approaches to motor control. Trends in Cognitive Sciences, 1(6), 209–216. https://doi.org/https://doi.org/10.1016/S1364-6613(97)01070-X

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. Science, 269(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Won, S. Y., Berger, J., & Slaney, M. (2014). Simulation of One's Own Voice in a Two-parameter Model. Proceedings of the International Conference on Music Perception and Cognition (ICMPC).

Woods, A., Jones, N., Alderson-Day, B., Callard, F., & Fernyhough, C. (2015). Experiences of hearing voices: Analysis of a novel phenomenological survey. The Lancet Psychiatry, 2(4), 323–331. https://doi.org/10.1016/S2215-0366(15)00006-1

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. Trends in Cognitive Sciences, 17(6), 263–271. https://doi.org/10.1016/j.tics.2013.04.004

Zarkali, A., Adams, R. A., Psarras, S., Leyland, L.-A., Rees, G., & Weil, R. S. (2019). Increased weighting on prior knowledge in Lewy body-associated visual hallucinations. Brain Communications, 1(1), fcz007. https://doi.org/10.1093/braincomms/fcz007

Zelano, C., Jiang, H., Zhou, G., Arora, N., Schuele, S., Rosenow, J., & Gottfried, J. A. (2016). Nasal Respiration Entrains Human Limbic Oscillations and Modulates Cognitive Function. The Journal of Neuroscience, 36(49), 12448–12467. https://doi.org/10.1523/JNEUROSCI.2586-16.2016

Abbreviations

AVH	Auditory-verbal hallucinations
BSC	Bodily self-consciousness
ECG	Electrocardiogram
EDA	Electrodermal activity
EEG	Electroencephalography
FA	False alarm
fMRI	Functional magnetic resonance imaging
FOVD	Familiar-other voice discrimination
GEV	Global Explained Variance
GFP	Global Explained Power
IFG	Inferior frontal gyrus
PET	Positron emission tomography
SCR	Skin conductance response
SoA	Sense of agency
SOVD	Self-other voice discrimination

Appendix

1

2

Enhancing analgesic neuromodulation through

personalized immersive virtual reality in patients with

chronic leg pain

6

7   Marco Solcà [1,2]†, M.D., PhD, Vibhor Krishna[3,4]†, M.D., Nicole Young[3,4], Ph.D.,

8   Milind Deogaonkar[6], M.D., Bruno Herbelin[1], Ph.D., Pavo Orepic[1], M.S., Robin

9   Mange[1],M.S., Giulio Rognini[1],Ph.D., Andrea Serino[1,5],Ph.D., Ali Rezai[6]‡,M.D.,

10 Olaf Blanke[1,7]‡ *,M.D., PhD

11

12 1   Laboratory of Cognitive Neuroscience, Center for Neuroprosthetics and Brain Mind
13   Institute, School of Life Sciences, Swiss Federal Institute of Technology (EPFL),
14   Lausanne, Switzerland
15 2   Department of Psychiatry, University Hospital, Geneva, Switzerland
16 3   Center for Neuromodulation, The Ohio State University, 480 Medical Center Dr,
17   Columbus, OH, 43210, USA
18 4   Department of Neurological Surgery, The Ohio State University, 410 W 10th Ave,
19   Columbus, OH, 43210, USA
20 5   MySpace, Department of Clinical Neurosciences, University Hospital of Lausanne,
21   Lausanne, Switzerland.
22 6   Rockefeller Neuroscience Institute, West Virginia University, Morgantown, West
23   Virginia
24 7   Department of Neurology, University Hospital, Geneva, Switzerland
25
26 †, ‡ : equal contribution
27

29

30 **Running head**: Neuromodulation-enhanced immersive virtual reality

31

32

33 *** Correspondence**:

34 Prof. Olaf Blanke
35 Laboratory of Cognitive Neuroscience

36 Center for Neuroprosthetics
37 Campus Biotech

38 Chemin des Mines 9

39 1202 Geneva

40 Switzerland
41 e-mail: olaf.blanke@epfl.ch

42

43

44

45

46

47 **Abstract**

48 Spinal cord stimulation (SCS) is an approved treatment for truncal and limb neuropathic pain.

49 However, pain relief is often suboptimal and SCS efficacy may reduce over time, sometimes

50 requiring addition of other pain therapies, stimulator revision, or even explantation. We

51 designed and tested a new procedure by combining SCS with immersive virtual reality (VR)

52 to enable analgesia in patients with chronic leg pain. We coupled SCS and VR by linking

53 SCS-induced paresthesia with personalized visual bodily feedback that was provided by VR

54 and matched to the spatio-temporal patterns of SCS-induced paresthesia. Here we show the

55 efficacy of neuromodulation-enhanced VR for the treatment of chronic pain by showing that

56 congruent SCS-VR reduced pain ratings on average by 44%. SCS-VR analgesia was stronger

57 than in two control conditions, kept increasing over successive stimulations, and persisted

58 after SCS-VR had stopped. Linking latest VR technology with recent insights from the

59 neuroscience of body perception and SCS-neuromodulation, our personalized new SCS-VR

60 platform highlights the impact of immersive digiceutical therapies for chronic pain.

61

62 **Registration**: clinicaltrials.gov Identifier: NCT02970006

63

64

65

66

67

68 **Introduction**

69 Chronic pain is a major health care problem affecting an estimated 20% of people worldwide,

70 accounting for approximately 40% of all medical visits, and costing more than $600 billion

71 annually in the United States alone[1]. Although several pharmacological treatments exist for

72 acute pain such as acetaminophen, nonsteroidal anti-inflammatory drugs (NSAIDS) and

73 opioids[2], current pharmacological treatments for chronic pain are often inadequate and are in

74 some cases associated with serious side effects such as physical dependence, tolerance, and

75 respiratory depression[3]. Accordingly, medical interventions such as physical or psychological

76 therapies, neurorehabilitation programs, or neuromodulation procedures (e.g. spinal cord or

77 brain stimulation, transcranial magnetic stimulation) have been developed to address

78 patients' needs and the enormous societal burden caused by chronic pain.

79

80 In particular, spinal cord stimulation (SCS) is widely used for the treatment for mixed

81 neuropathic-nociceptive and neuropathic-radicular pain from conditions such as failed back

82 surgery syndrome (FBSS) and complex regional pain syndrome (CRPS)[4]. SCS-induced

83 analgesia is mediated by several mechanisms[5], but generally linked with SCS-induced

84 paresthesias that arise from the electrical stimulation of Aβ fibers masking painful sensations

85 by more comfortable tingling sensations[6]. SCS substantially reduces pain in about half of the

86 patients, but even when efficient, rarely eliminates pain completely[7]. In addition, the

87 analgesic effects of SCS-induced parasthesias have been reported to decrease over time[4],

88 requiring revision and explantation of the implanted electrode array in approximately 25% of

89 cases[8]. Additional oral medications have not shown additional and persistent benefits on SCS

90 outcomes[9], except add-on therapy with intrathecal analgesic delivery or subcutaneous

91 stimulation, which have been proposed to overcome some of these limitations and boost the

92 analgesic effects of SCS[10,11].

93

94   Next to pharmacological and neuromodulation pain therapy, there has been a major upsurge

95   in digital technologies that were initially developed for very different fields such as media,

96   entertainment, and research and have recently been adapted as non-invasive treatments for

97   chronic pain. Following early work of Hoffman and colleagues[12], gaming elements have been

98   inserted into immersive virtual reality (VR) environments and demonstrated to have analgesic

99   properties based on cognitive-distractive effects in patients with severe skin burns[13]. More

100   recent work inspired by research in the cognitive neuroscience of embodiment, integrated the

101   patient's body or avatar into the virtual scene (shown on a head-mounted display, HMD) and

102   further leveraged VR to provide personalized and automatized multisensory bodily inputs.

103   Tapping into central body representations and inducing a range of well-controlled illusory

104   bodily experiences, these authors induced illusory perceptions of the hand[14], leg[15], face[16], and

105   body-torso[17]. Thus, in the virtual leg illusion[15] participants feel illusory touch and ownership

106   over virtual legs that they observe in VR, where these legs are seen as being touched at the

107   same time and place as their own physical legs. Importantly, such multisensory stimulations

108   have demonstrated analgesic properties by showing that visuo-tactile[18] or visuo-motor[19]

109   stimulation (applied manually) can relieve patients' chronic pain.

110

111   However, the delivery of visuo-tactile and visuo-motor stimulation are not always possible

112   and may be counterproductive, because in many patients with chronic pain movement and

113   even gentle touch to the affected limb may increase or induce pain (i.e. allodynia). Moreover,

114   classical multisensory approaches are not automatized and require the patient's (or

115   therapist's) active participation reducing the possibility of prolonged and/or home-based

116   treatment. To overcome these limitations, we have recently designed fully automatized

117   stimulation techniques that are integrated into our digiceutical VR platform and based on

118   either additional interoceptive cues such as the heartbeat[20–22], based on additional peripheral

119   nerve stimulation (inducing paresthesia), or transcranial magnetic stimulation[23–25] with

120   analgesic efficacy in chronic pain[20,24].

121

122   We here extended our digiceutical platform by combining two distinct pain therapies,

123   adapting them for the first time to 15 patients suffering from severe chronic leg pain with a

124   spinal cord implant, and investigated its analgesic effects. In this novel SCS-VR solution,

125   epidural SCS provides neurotactile stimulation leading to paresthesia in the painful body part

126   while the VR environment is programmed to add personalized visual feedback to

127   superimpose a virtual pattern in the location where the patient feels the SCS-induced

128   paresthesias. Thus, our digiceutical platform allowed us to tailor the procedure to each

129   patient's pain characteristics, providing multisensory bodily stimulation with specific spatio-

130   temporal stimulation patterns, while avoiding the application of potentially harmful physical

131   touch cues (i.e. avoiding allodynia) to the affected body part. In the present report, we

132   describe the method and procedure for the first time and test the hypothesis that our SCS-VR

133   digiceutical method boosts analgesia (versus incongruent SCS-VR and VR alone control

134   conditions; see below). Our new immersive VR platform allowed us to record and immerse

135   patients in any 360° stereoscopic video environment (Fig. 1A & 1B) and to use real-time

136   integration of the patient's own body in the 360° scene, as if seen from their first-person

137   perspective via a head-mounted display (HMD) (Fig. 1C). In three different conditions we

138   provided online visual illumination of a circumscribed skin region on the patients' legs seen

139   in VR corresponding to the onset and somatotopic location of circumscribed touch sensations

140   on the patient's leg as elicited by SCS (Fig. 1D).

141
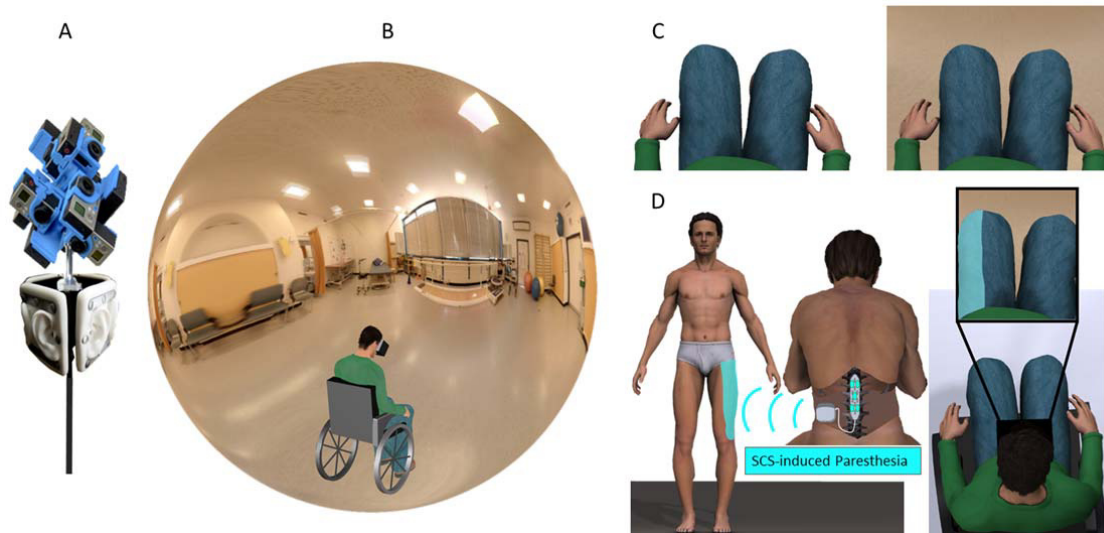
142

143 **Figure 1** <u>VR enhanced SCS</u>: (A) The experimental virtual scenario was created by capturing a real
144 environment using a panoramic and stereoscopic device composed of 14 go-pro cameras and 8
145 microphones located inside ear-shaped silicon molds. (B) Patients were immersed in this pre-
146 recorded environment using a VR headset (Oculus Rift CV1; 2160 x 1200 per eye, 110° FOV) and
147 RealiSM software (http://lnco.epfl.ch/realism). (C) Infrared stereoscopic camera, placed on the
148 head-mounted display (HMD), captured patient's own body in real time, which was merged in the
149 prerecorded scene. (D) The SCS-VR platform allowed to integrate and visualize SCS-induced
150 paresthesia.

151

152 We measured analgesic effects of SCS-VR (i.e. Congruent SCS-VR condition) in all 15

153 patients and tested immediate analgesic effects (several times during SCS-VR exposure) and

154 carry over effects (10 minutes after SCS-VR had ended). These data were compared to those

155 from two well-matched control conditions designed to control for (1) potential analgesic

156 effects of seeing one's own body in VR without SCS stimulation (VR alone) and for (2)

157 spatio-temporal congruency between the visual illumination and the SCS-induced tactile

158 sensations (Incongruent SCS-VR condition; i.e. same VR and SCS stimulation but misaligned

159 in space) (Fig. 2).

Congruent SCS-VR   Incongruent SCS-VR   VR alone

160
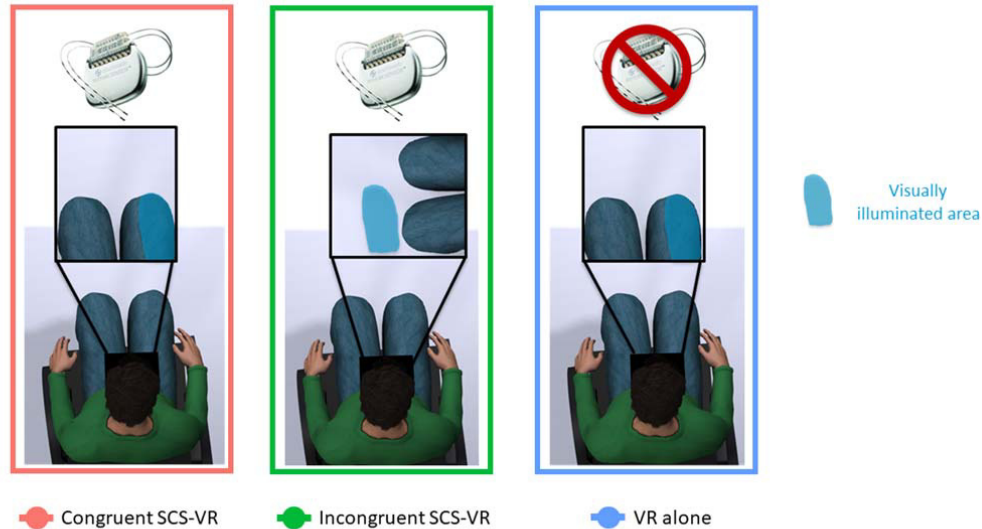
Figure 2   <u>Experimental condition:</u> Each implanted SCS patients underwent three different experimental conditions. During Congruent SCS-VR condition, visual VR feedback was superimposed at the location where each patient felt SCS-induced paresthesias on his/her body (as defined during prior personalization session), as seen from a first-person visual perspective (Red). In the Incongruent SCS-VR condition, the same visual feedback was applied but the virtual body and legs were tilted by 90° thereby inducing spatial misalignment between visual VR feedback and SCS-induced paresthesia (Green). During the VR alone condition patients observed visual illumination of their body (as seen during Congruent SCS-VR condition) but in absence of SCS-induced paresthesia (i.e. stimulator switched off) (Blue).

Results

All participants tolerated the entire procedure well and none reported adverse effects related to the experiment.

Group-level SCS-VR analgesia

176   Figure 3A reports the evolution of pain ratings from baseline during stimulation for the three

177   experimental conditions and shows significant pain reduction following Congruent SCS-VR

178   (i.e. pain rating at the end of the stimulation significantly different from baseline) (t(14)=-

179   4.11, p=0.001) and a pain reduction approaching significance following Incongruent SCS-VR

180   condition (t(14)=-1.989, p=0.067). Conversely, no significant analgesia was observed

181   following the VR alone condition (t(14)=0.26, p=0.798). Further statistical analysis (linear

182   mixed effects model; fixed effect of visual feedback, time and their interaction, random

183   intercepts for subjects as well as by-subject random slope for the effect of condition) revealed

184   a significant interaction between time and experimental condition (F (2, 270) = 11.10, p <

185   0.001). As predicted, post-hoc pairwise comparisons of the post-stimulation pain values

186   revealed a significantly larger analgesic effect during the Congruent SCS-VR condition than

187   the two control conditions (i.e. VR alone: t(14)=3.56, p=0.003; Incongruent SCS-VR:

188   t(14)=3.74, p=0.002). In addition, a difference approaching significance between Incongruent

189   SCS-VR and VR alone was observed (t(14)=1.97, p=0.071). Differences in baseline pain

190   ratings (i.e. pain rating before SCS-VR was started) were not significantly different between

191   the three conditions (F(2)=0.636, p=0.537). Collectively, these results support the idea that

192   SCS-VR enhances SCS-induced analgesia when the multisensory signals are congruently

193   presented, that is when the tactile sensations provided through SCS and the illuminated skin

194   regions are aligned in space.
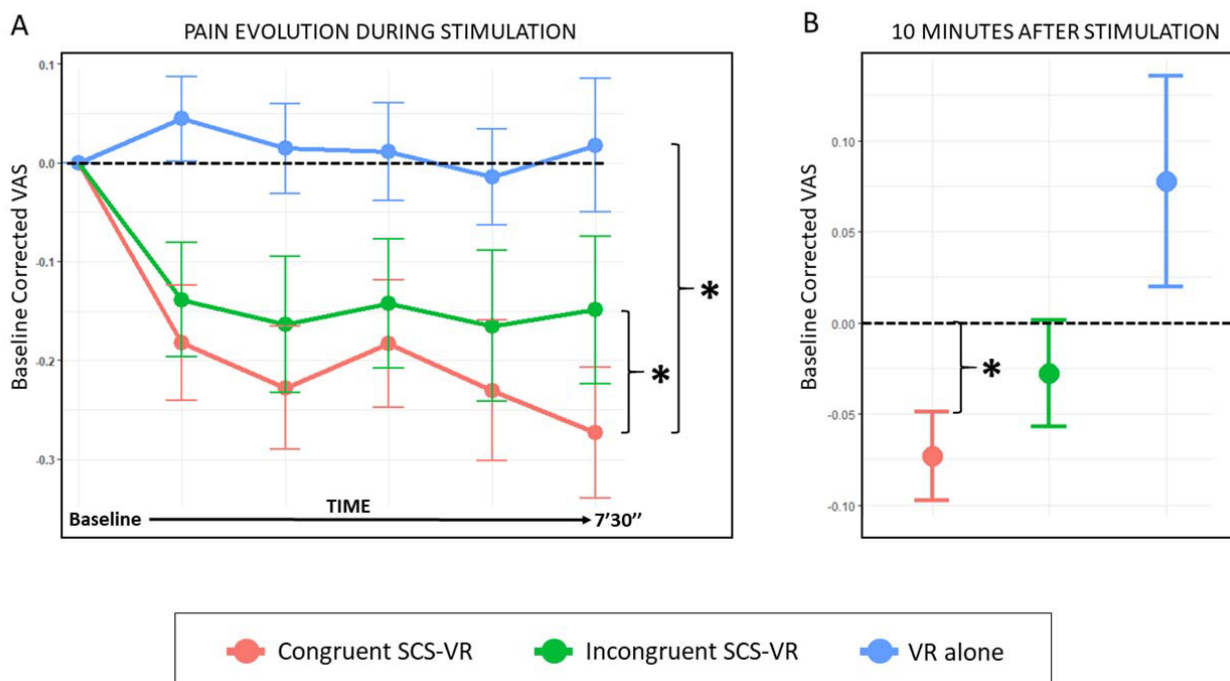
Figure 3   <u>Analgesic effects</u>: (A)   Pain ratings during and (B) 10 minutes after SCS-VR had ended are shown in the Congruent SCS-VR (Red), the Incongruent SCS-VR (Green) and the VR alone conditions (Blue). *p < 0.05.

Effect size of SCS-VR analgesia

On average, pain levels were reduced by 2.72 (95% CI 4.15 to 1.31) (from a maximum rating of 10 on the VAS) during the Congruent SCS-VR condition, which represents a reduction of 44% (95% CI -60 to -30) from the baseline pain value (average baseline pain rating = 6.2, 95% CI 4.8 to 7.6). This reduction during congruent SCS-VR was on average 1.8 times larger in magnitude than analgesia reported when SCS and visual feedback were misaligned (i.e. Incongruent SCS-VR, 23% pain reduction, 95% CI -46 to -0.3)) and even more compared to VR alone (3% pain reduction, 95% CI -24 to 17). Modelling the pain ratings in the Congruent SCS-VR condition as function of time, we found a significant negative slope for pain rating over time ($F_{(7.4)}$=-2.63, p=0.01), estimating the pain reduction per minute at

211   0.07 out of 10 (SEM = 0.006). No such effects of slope were observed in the other two

212   conditions (Incongruent SCS-VR: $F(7.4)=-0.49$, $p=0.62$, VR alone: $F(7.4)=1.23$, $p=0.21$),

213   suggesting that the more often congruent SCS-VR was applied, but not congruent SCS-VR or

214   VR alone, the larger was its analgesic effect.

215

216   *Analysis of individual data*

217   Investigating analgesia for individual patients, we found that SCS-VR led in 14 out of the

218   tested 15 participants to a pain reduction with respect to baseline in the Congruent SCS-VR

219   condition (i.e. negative baseline corrected values at the end of the stimulation). Moreover, the

220   same number of patients had stronger analgesia during the Congruent SCS-VR condition

221   compared to Incongruent SCS-VR condition. A binomial test indicated that the proportion of

222   patients with stronger analgesic during SCS-VR compared to control conditions was larger

223   than chance level ($p < .001$, 2-sided). Of note, four patients, differently from the rest of the

224   group, received a high-frequency stimulation pattern (Table 1), which provides paresthesia-

225   free pain relief. Pain relief was equivalent in patients with high frequency and conventional

226   SCS, suggesting that SCS-VR can be used with either conventional (40-60 Hz) or high

227   frequency protocols (800-1000 Hz). Although the high frequency stimulation does not induce

228   conscious tactile perceptions (paresthesias), the stimulation is programmed with electrodes

229   'covering' the sensory fibers of the specific body region where the patient experiences pain.

230

231   *Persisting analgesia*

232   Figure 3B shows the pain ratings for the three different conditions, measured 10 minutes after

233   SCS-VR had ended. Analgesia was only observed after Congruent SCS-VR condition (one

234   sample t-test $t(14)=-3.02$, $p=0.009$), whereas no such lasting effect was observed following

235   the Incongruent SCS-VR condition (one sample t-test t(14)=-0.95, p=0.36,) or the VR alone

236   condition (one sample t-test t(14)=1.34, p=0.19).

237

238   *Embodiment*

239   Embodiment as assessed through questionnaire ratings is shown in Figure 4. The first two

240   questions investigated each patient's subjective sensation related to the SCS-VR, that is

241   ownership for the virtual legs (the feeling that the seen legs were part of their body) and

242   illusory touch (the feeling that the illumination caused the SCS-induced paresthesia), while a

243   third question, unrelated to the illusion, served as control for suggestibility. As predicted, we

244   observed a main effect of questions (F(2)=19.93, p=0.0004) and post-hoc pairwise

245   comparisons revealed a significant difference between ownership and illusory touch versus

246   the control item (*illusory touch* vs. *control*: (t(14)=-2.60, p=0.02); (*ownership* vs. *control*:

247   (t(14)=-6.11, p<0.001), confirming that VR-SCS induced changes in leg embodiment. No

248   other effects were significant (all p > 0.56).
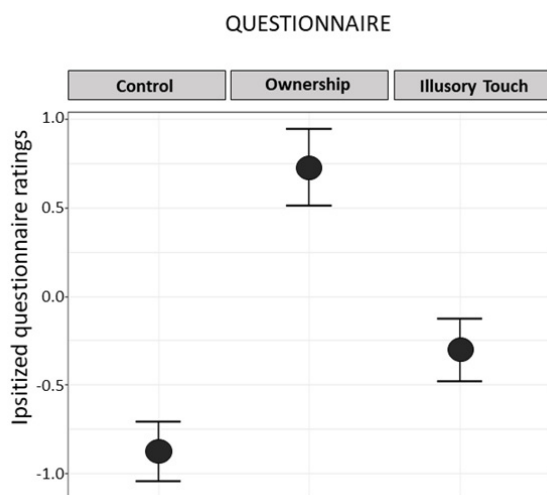
249

250   **Figure 4 :**   Embodiment: 3-items questionnaire investigating self-identification and referred touch shows

251   that VR-SCS induced changes in leg embodiment (see main text for details)

252

253   **Discussion**

254   We describe and apply a new digiceutical method for chronic pain, SCS-VR, that combines

255   neuromodulation, VR, and latest research from cognitive neuroscience of multisensory

256   integration into a single therapeutic solution. Based on a recently described digiceutical

257   technique that integrates VR-based visual stimulation with neural stimulation of

258   intrafascicular implants for research in upper limb amputees[24], we here used a similar concept

259   and integrated VR-based visual stimulation with neural stimulation of spinal cord implants,

260   which is a widely used treatment in patients with chronic leg/back pain, a frequent pain

261   disorder. Compared to existing other add-on therapies for SCS such as intrathecal anaelgesic

262   delivery or subcutaneous stimulation[10,11], SCS-VR boosts SCS-related effects in a completely

263   non-invasive fashion and avoids the risk of allodynia that many chronic pain patients report

264   even during the therapeutic application of mild touch cues or limb movements during

265   physical therapy. The present SCS-VR protocol may thereby facilitate the application of

266   prolonged and more frequent therapy sessions in more patients suffering from chronic pain.

267

268   SCS-VR had a stronger analgesic effect than VR alone, which was a condition that consisted

269   of congruent VR, but without any SCS. Thus, any potential distractive effects of VR per se or

270   any analgesic effects due to the mere viewing of a body in VR[34] cannot explain the present

271   SCS-VR pain relief. Similarly, analgesia observed during SCS-VR cannot be attributed to

272   SCS alone, as the effect in the Congruent SCS-VR condition was 1.8 times stronger than in

273   the Incongruent SCS-VR condition, when VR and SCS stimulations were both present, but

274   not aligned in space. Thus, it is not sufficient to expose patients to VR or to add a view of a

275   patient's legs or body in VR to SCS: the observed SCS-VR analgesic effect is due to the

276   combination of SCS-induced paresthesia and congruent illumination of the circumscribed

277   skin region on the patient's virtual leg that corresponds to the somatotopic location of SCS-

278   induced paresthesia. Therefore, this study shows, for the first time, that it is possible to

279   integrate immersive and multisensory VR with spinal neuromodulation to increase the

280   analgesic effects of SCS. Despite evidence for several mechanisms underlying SCS-induced

281   analgesia[5], the main hypothesis posits that noxious afferent activity is successfully masked or

282   inhibited by non-noxious paresthesias that are induced by the stimulation of the dorsal

283   column of the lumbar spinal cord segment[6]. This link between tactile processing and SCS-

284   induced analgesia and the present SCS-VR effects are also supported by clinical observations

285   revealing that the level of overlap between the patient's location of pain and the location of

286   SCS-induced paresthesia is an important predictor of SCS therapeutic outcome[35].

287   Accordingly, we propose that the analgesic effects of our novel SCS-VR setup is likely based

288   on enhanced masking of nociceptive information through congruent visual and tactile signals.

289   The present data, and, in particular, the significant differential analgesic effects in the

290   congruent SCS-VR versus incongruent SCS-VR conditions lend further support to this

291   proposal and add the novel finding that such bodily overlap should not only involve the site

292   of pain and paraesthesia, but also visual bodily cues in VR. Further supported by data from

293   our two carefully designed control conditions, we suggest that SCS-VR anagelsia is primarily

294   based on cortical effects that subsequently modulate subcortical and spinal pain processing

295

296   In the four patients, receiving high-frequency SCS that did not induce paresthesias, we

297   observed pain relief that was equivalent to patients receiving conventional SCS. This suggests

298   that SCS-VR can be used with either conventional (40-60 Hz) or high frequency protocols

299   (800-1000 Hz) and in the latter case without exposing the patient to conscious tactile

300   perceptions (paresthesias). Importantly, basic research in multisensory perception has shown

301   that visuo-tactile integration occurs even when external or body-related stimuli are presented

302   without conscious perception (e.g.[36,37]), and multisensory interactions may even be stronger

303 in conditions with lower (vs. higher) multisensory stimulations[38,39]. Thus, it is reasonable to

304 argue that SCS-VR enhances cortical-spinal somatosensory processing even when not

305 consciously perceived and therefore can be equally efficient to boost analgesia in high-

306 frequency SCS protocols and may facilitate the application of more frequent SCS-VR dosage.

307 However, this has to be regarded with caution, as these data need to be confirmed in a larger

308 cohort of patients.

309

310

311 *Limitations*

312 We note that, although we carefully controlled the multisensory and SCS aspects of SCS-VR,

313 the three experimental conditions were easily distinguishable and therefore a blinded design

314 was not feasible (even though all patients were naïve about the aim of the different

315 experimental condition). Another limitation was our inability to fully randomize the

316 experimental conditions; indeed, in order to ensure the absence of any persisting SCS

317 analgesic effects during the VR alone condition, each patient was asked to switch off his/her

318 stimulator the day before the experimental session (wash out) and we always started with the

319 VR alone condition (see method section). However, the order of the two key VR conditions

320 to be compared, the Congruent and Incongruent SCS-VR conditions, were carefully

321 counterbalanced between participants, excluding an effect due to order of conditions.

322

323 *Clinical relevance*

324 Pain reduction was selective and reliable - as patients repeatedly experienced less pain during

325 Congruent SCS-VR stimulation and this reduction increased over time and all different short

326 therapy sessions. This suggests that longer and more repeated stimulation increases SCS-VR

327   analgesia, indicating its potential clinical relevance. Moreover, SCS-VR reduced chronic leg

328   pain in 14 out of 15 patients with an average of 44% pain reduction after only 7.5 minutes of

329   stimulation. According to IMMPACT recommendations[40], this corresponds to clinically

330   meaningful analgesia, although the applied SCS-VR exposure was much shorter than other

331   pain therapies(4)[5]. Finally, the analgesic effect outlasted SCS-VR, persisting at least ten

332   minutes after the end of SCS. No such lasting analgesia was observed in any control

333   condition suggesting that SCS-VR may have more enduring effects than conventional

334   approaches.

335

336   In conclusion, we achieved SCS-VR analgesic effects through a fully automatized stimulation

337   that avoids the application of potentially painful bodily cues, while minimizing the active

338   involvement of patient and therapist. We argue that the strength of the effect, its selectivity,

339   its ease of application, and consistent increase across sessions and long-term analgesia will

340   facilitate the application of prolonged and more frequent therapy doses in future SCS-VR

341   studies, likely further boosting the described effects. Another advantage of the present

342   digiceutical platform is the possibility to personalize and to gamify the SCS-VR therapy as

343   well as translate the SCS-VR setting to the home of patients which may improve patients'

344   motivation and compliance with the treatment, thus potentially counteracting the reduction of

345   SCS effectiveness over time[4] and the need of explanting the electrodes[9].

346

347   **Materials and Methods**

348   *Study Design*

349   This study was registered at ClinicalTrials.gov (NCT02970006) and was designed to test the

350   hypothesis that SCS-induced analgesia could be enhanced with an immersive digital system

351   enabling visual representation of SCS-induced paresthesia. To this aim, we developed an

352   immersive VR platform and exposed SCS implanted patients suffering from chronic leg pain

353   with a patient-tailored virtual scenario where they observed a visual illumination pattern

354   corresponding to the sensations experienced on their real legs as induced by SCS. We

355   assessed pain rating as primary outcome and embodiment (assessed through questionnaire) as

356   secondary measure (see below). Based on Pozeg and collaborators[18] we estimated that a

357   sample size of 15 patients was needed to observe a significant analgesic effect considering a

358   power of test = 90% and a significant level =0.05. Patient recruitment stopped when 15

359   patients were included.

360

361   *Participants*

362   21 patients with spinal cord stimulator implant for chronic leg pain were screened and

363   consented from the Center for Neuromodulation at The Ohio State University Wexner

364   Medical Center between October 1, 2017 and January 31, 2019. 15 patients (5 women, mean

365   age: 47.7 years; SD: ±9.56 range: 33–61 years, mean time since surgery: 5.3 months SD:

366   ±2.2) fulfilled inclusion criteria (Table 2) and accepted to participate in our study. All patients

367   had normal or corrected to normal vision and were naïve about the manipulation (control-

368   experimental conditions). Approval for this study was obtained from the Ohio State

369   University Medical Center Institutional Review Board (Columbus, Ohio). All patients

370   completed an informed consent process before their participation in the study. Patient's

371   demographic and clinical characteristics are listed in Table 1.

372

373   *Procedure*

374   The experiment included two sessions occurring at 24 hours interval.

375   1) <u>Clinical assessment and VR setting session</u>

376   Initial clinical assessment included medical interview concerning the patient's history, the

377   underlying pain diagnosis and the impact of pain on patient's daily activities scored with the

378   Oswestry Disability Index [26]. Then, visual illumination of a circumscribed skin region on the

379   patient's virtual legs was personalized individually to correspond to the tactile sensations

380   induced by the SCS . To this aim, patients were asked to look at their own body while

381   wearing the VR headset. First, the experimenter, guided by the patient's feedback, indicated

382   the area of the body where tactile sensation (or pain for patients with high-frequency settings)

383   was experienced. Then, the defined region was illuminated by manipulating visual parameters

384   such as color intensity, size of the visual pattern and frequency of illumination until it best-

385   matched patient's subjective feeling. At each step, patients gave direct feedback on the VR

386   parameters. Once defined, the optimal illumination parameter tailored for each patient was

387   stored and reloaded during the experimental session. For patients with high-frequency

388   settings, standard parameters were used based on the pilot experiment. This first session

389   lasted approximately 1.5 hours.

390

391   2) <u>Experimental session</u>

392   Patients were instructed to switch their stimulator off the night before coming for the second

393   session for wash out of SCS related pain relief. Each patient started with the VR alone

394   condition (to ensure the absence of potential long-lasting effect of SCS) during which patients

395   observed visual illumination of their body (as defined during the first session) but in the

396   absence of SCS-induced paresthesia (i.e. stimulator switched off). Then, each patient

397   underwent the other two conditions (randomized across participant) where the stimulator was

398   switched on. During Congruent SCS-VR stimulation, visual feedback was provided on

399   patient body, as seen from   a first-person perspective. In the Incongruent SCS-VR condition,

400   the body was tilted by 90° to induce a misalignment between the visual feedback and SCS-

401   induced paresthesia (Figure 2). Each condition started outside the virtual environment with

402   the current subjective pain measure. Then the subject was assisted in putting on the VR

403   headset (adjusted and calibrated for stereoscopic vision for each participant) and instructed to

404   look at his own legs, seen through the head mounted display (HMD). Stimulation (VR

405   illumination and/or SCS) lasted for 7'30'' intermingled with pain ratings every 90 seconds

406   (see below). At the end of the stimulation, VR headset was removed and/or SCS switched off

407   and patients answered to a questionnaire. 10 minutes after the end of the simulation, patients

408   were asked again to assess their ongoing pain (long-term value). Between conditions, subjects

409   had a 15 minutes break. This entire experimental session lasted approximately 2 hours.

410

411

412   *Immersive digital platform*

413   To test the hypothesis that a visual representation of SCS-induced paresthesia would boost

414   SCS-induced analgesia, we developed an immersive digital scenario with an in-house

415   software developed at the Blanke Lab. This technology (the RealiSM software, Reality

416   Substitution Machine, http://lnco.epfl.ch/realism) constitutes a new approach to virtual reality

417   immersion, body integration and stimulus presentation for cognitive neurosciences. The

418   experimental scene was built as follows. First, we captured a real environment using a

419   panoramic and stereoscopic device composed of 14 go-pro cameras (GoPro Hero3+).

420   Binaural Sound from the environment was also recorded by 8 microphones located inside ear-

421   shaped silicon molds (Figure 1A).

422   Using this technology, we recorded for the present study a neutral environment

423   (physiotherapy room) where patients were immersed using an Oculus Rift VR headset (2160

424   x 1200 display resolution; 110° field of view, refreshing rate 90Hz, head orientation tracking

425   through accelerometer, gyroscope, magnetometer and Constellation tracking camera) (Figure

426   1B). Infrared stereoscopic camera (DUO MLX R2, http://duo3d.com/docs/articles/duo-mlx),

427   placed on the Oculus headset allowed body extraction (in black and white) which was

428   integrated in real time within the previously recorded environments using the RealiSM

429   platform. An additional webcam (Logitech HD Webcam C310 960-000588) was attached to

430   the Oculus to extract the color of patient's body and integrated into the stereoscopic image of

431   the body obtained with the infrared camera (Figure 1C). Thus, a real-time color video of the

432   subject's body was merged with pre-recorded, tridimensional environment creating a high

433   degree of immersion through instant visual feedback of patient's body movements.

434   This technology also allowed us to provide additional visual feedback through highly

435   controlled and realistic experimental manipulations of the virtual scenes. Linking for the first

436   time neuromodulation and VR, we illuminated the online recorded real legs of the

437   participants accordingly to the sensations they experienced as induced by the SCS (Figure

438   1D). Illumination was created alternating 10 different textures composed of dots of various

439   blue intensity. The frequency of texture presentation, dots' size and color intensity was

440   adapted during the first session based on patients' report to correspond to the tactile sensation

441   elicited by the SCS stimulation.

442

443   *Measurements*

444   Analgesic effects

445   Subjective pain perception was measured before starting each condition (i.e. baseline

446   measure), every 90 sec during neuro-visual stimulation, at the end of the stimulation (7.5

447   minutes), and 10 minutes after stimulation (long-term). We asked participants to place a mark

448   at the point that represents their actual pain intensity using a vertical visual continuous analog

449   scale (VAS), ranging from "no pain" (bottom of scale) to "the worst imaginable pain" (top of

450   scale). During stimulation, the vertical visual scale was presented virtually and patients

451   moved a marker using a wireless mouse. Difference between each rating and the baseline pain

452   measure was used for data analysis (baseline correction).

453

454   Questionnaire

455   After each condition, we administered a 3-items questionnaire adapted from Pozeg and

456   collaborators[18] to investigate ownership («I had the impression that the legs I was looking at

457   were my real legs») and illusory touch («I had the impression that the illumination was

458   causing the tingling sensation») and a control item («I had the impression that my legs

459   disappeared»). Participants were asked to indicate how much they agreed with each item

460   using a 7-point colored vertical Likert scale ranging from 0 (complete disagreement, the

461   bottom extreme, red point) to +6 (complete agreement, the top extreme, green point)

462

463   *Statistical analysis*

464   Analysis was performed with the software R (R Development Core Team, 2013) including

465   the "lme4" package[27]. We analyzed pain VAS using mixed model and minimization of

466   Bayesian Information Criterion (BIC)[28] was used for model selection (selected model: fixed

467   effect of visual feedback, time and their interaction, random intercepts for subjects as well as

468   by-subject random slope for the effect of condition). P-values were obtained by likelihood

469   ratio tests, and degrees of freedom were estimated with Satterthwaite approximation [29].

470   Questionnaire ratings (which were not normally distributed) underwent an intra-subject

471   standardization by means of an ipsatization procedure leading to normally distributed Z-

472   scores values allowing the use of parametric tests [30–33]. We then conducted 3 x 3 repeated

473   measures ANOVA with the factors questions (3 items), conditions (3 conditions).

474

475   *Data availability*

476   Raw data and the custom-developed VR software will be made available by materials transfer

477   agreement upon reasonable request to OB.

478

484

485   **Author contributions**

486   MS, VK, GR, AS, AL & OB designed the study. MS, VK, RM, PO, TG, NY & BH

487   developed the methodology and collected the data. MS, VK, AR and OB performed the

488   analysis, and interpreted the results, MS, VK, AR and OB wrote the initial draft of the paper

489   and all authors provided critical review, edits, and approval for the final manuscript.

490

491   **Competing interests**

492 The authors declare that they have no competing interests

493

494

495 **References**

496 1. Breivik, H., Collett, B., Ventafridda, V., Cohen, R. & Gallacher, D. Survey of chronic

497 pain in Europe: prevalence, impact on daily life, and treatment. *Eur. J. Pain Lond. Engl.*

498 **10**, 287–333 (2006).

499 2. Moore, R. A., Derry, S., Aldington, D. & Wiffen, P. J. Single dose oral analgesics for

500 acute postoperative pain in adults - an overview of Cochrane reviews. in *Cochrane*

501 *Database of Systematic Reviews* (John Wiley & Sons, Ltd, 2015).

502 3. Birnbaum, H. G. *et al.* Societal Costs of Prescription Opioid Abuse, Dependence, and

503 Misuse in the United States. *Pain Med.* **12**, 657–667 (2011).

504 4. Turner, J. A., Loeser, J. D., Deyo, R. A. & Sanders, S. B. Spinal cord stimulation for

505 patients with failed back surgery syndrome or complex regional pain syndrome: a

506 systematic review of effectiveness and complications. *Pain* **108**, 137–147 (2004).

507 5. Oakley, J. C. & Prager, J. P. Spinal Cord Stimulation: Mechanisms of Action. *Spine* **27**,

508 2574 (2002).

509 6. Zhang, T. C., Janik, J. J. & Grill, W. M. Mechanisms and models of spinal cord

510 stimulation for the treatment of neuropathic pain. *Brain Res.* **1569**, 19–31 (2014).

511 7. Taylor, R. S., Desai, M. J., Rigoard, P. & Taylor, R. J. Predictors of Pain Relief

512 Following Spinal Cord Stimulation in Chronic Back and Leg Pain and Failed Back

513 Surgery Syndrome: A Systematic Review and Meta-Regression Analysis. *Pain Pract.* **14**,

514 489–505 (2014).

515   8. Hayek, S. M., Veizi, E. & Hanes, M. Treatment-Limiting Complications of Percutaneous

516   Spinal Cord Stimulator Implants: A Review of Eight Years of Experience From an

517   Academic Center Database. *Neuromodulation J. Int. Neuromodulation Soc.* **18**, 603–608;

518   discussion 608-609 (2015).

519   9. Maher, D. P. *et al.* Neuropathic Pain Medication Use Does Not Alter Outcomes of Spinal

520   Cord Stimulation for Lower Extremity Pain. *Neuromodulation Technol. Neural Interface*

521   **21**, 106–113 (2018).

522   10. van Gorp, E.-J. J. A. A. *et al.* Subcutaneous Stimulation as ADD-ON Therapy to Spinal

523   Cord Stimulation Is Effective in Treating Low Back Pain in Patients With Failed Back

524   Surgery Syndrome: A Multicenter Randomized Controlled Trial. *Neuromodulation J. Int.*

525   *Neuromodulation Soc.* **19**, 171–178 (2016).

526   11. Goto, S. *et al.* Spinal cord stimulation and intrathecal baclofen therapy: combined

527   neuromodulation for treatment of advanced complex regional pain syndrome. *Stereotact.*

528   *Funct. Neurosurg.* **91**, 386–391 (2013).

529   12. Hoffman, H. G., Doctor, J. N., Patterson, D. R., Carrougher, G. J. & Furness, T. A.

530   Virtual reality as an adjunctive pain control during burn wound care in adolescent

531   patients. *Pain* **85**, 305–309 (2000).

532   13. Morris, L. D., Louw, Q. A. & Grimmer-Somers, K. The effectiveness of virtual reality on

533   reducing pain and anxiety in burn injury patients: a systematic review. *Clin. J. Pain* **25**,

534   815–826 (2009).

535   14. Botvinick, M. & Cohen, J. Rubber hands 'feel' touch that eyes see. *Nature* **391**, 756–756

536   (1998).

537   15. Pozeg, P., Galli, G. & Blanke, O. Those are Your Legs: The Effect of Visuo-Spatial

538   Viewpoint on Visuo-Tactile Integration and Body Ownership. *Front. Psychol.* **6**, 1749

539   (2015).

540   16. Tsakiris, M. Looking for myself: current multisensory input alters self-face recognition.

541   *PloS One* **3**, e4040 (2008).

542   17. Lenggenhager, B., Tadi, T., Metzinger, T. & Blanke, O. Video ergo sum: manipulating

543   bodily self-consciousness. *Science* **317**, 1096–1099 (2007).

544   18. Pozeg, P. *et al.* Virtual reality improves embodiment and neuropathic pain caused by

545   spinal cord injury. *Neurology* (2017) doi:10.1212/WNL.0000000000004585.

546   19. Ortiz-Catalan, M. *et al.* Phantom motor execution facilitated by machine learning and

547   augmented reality as treatment for phantom limb pain: a single group, clinical trial in

548   patients with chronic intractable phantom limb pain. *The Lancet* **388**, 2885–2894 (2016).

549   20. Solcà, M. *et al.* Heartbeat-enhanced immersive virtual reality to treat complex regional

550   pain syndrome. *Neurology* **91**, e479–e489 (2018).

551   21. Park, H.-D. *et al.* Neural Sources and Underlying Mechanisms of Neural Responses to

552   Heartbeats, and their Role in Bodily Self-consciousness: An Intracranial EEG Study.

553   *Cereb. Cortex N. Y. N 1991* **28**, 2351–2364 (2018).

554   22. Park, H.-D. & Blanke, O. Coupling Inner and Outer Body for Self-Consciousness. *Trends*

555   *Cogn. Sci.* **23**, 377–388 (2019).

556   23. Bassolino, M. *et al.* Non-invasive brain stimulation of motor cortex induces embodiment

557   when integrated with virtual reality feedback. *Eur. J. Neurosci.* **47**, 790–799 (2018).

558   24. Rognini, G. *et al.* Multisensory bionic limb to achieve prosthesis embodiment and reduce

559   distorted phantom limb perceptions. *J Neurol Neurosurg Psychiatry* jnnp-2018-318570

560   (2018) doi:10.1136/jnnp-2018-318570.

561   25. Franza, M. *et al.* Hand perceptions induced by single pulse transcranial magnetic

562   stimulation over the primary motor cortex. *Brain Stimulat.* **12**, 693–701 (2019).

563   26. Fairbank, J. C. & Pynsent, P. B. The Oswestry Disability Index. *Spine* **25**, 2940–2952;

564   discussion 2952 (2000).

565   27. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models

566   Using lme4 | Bates | Journal of Statistical Software.

567   https://www.jstatsoft.org/article/view/v067i01.

568   28. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **6**, 461–464 (1978).

569   29. Satterthwaite, F. E. An approximate distribution of estimates of variance components.

570   *Biometrics* **2**, 110–114 (1946).

571   30. Cattell, R. B. Psychological measurement: normative, ipsative, interactive. *Psychol. Rev.*

572   **51**, 292–303 (1944).

573   31. Broughton, R. & Wasel, N. A text-stimuli presentation manager for the IBM PC with

574   ipsatization correction for response sets and reaction times. *Behav. Res. Methods Instrum.*

575   *Comput.* **22**, 421–423 (1990).

576   32. Romano, D., Pfeiffer, C., Maravita, A. & Blanke, O. Illusory self-identification with an

577   avatar reduces arousal responses to painful stimuli. *Behav. Brain Res.* **261**, 275–281

578   (2014).

579   33. Ronchi, R. *et al.* Right insular damage decreases heartbeat awareness and alters cardio-

580   visual effects on bodily self-consciousness. *Neuropsychologia* **70**, 11–20 (2015).

581   34. Longo, M. R., Betti, V., Aglioti, S. M. & Haggard, P. Visually Induced Analgesia: Seeing

582   the Body Reduces Pain. *J. Neurosci.* **29**, 12125–12130 (2009).

583   35. North, R. B., Ewend, M. G., Lawton, M. T. & Piantadosi, S. Spinal cord stimulation for

584   chronic, intractable pain: superiority of 'multi-channel' devices. *Pain* **44**, 119–130

585   (1991).

586   36. Faivre, N., Mudrik, L., Schwartz, N. & Koch, C. Multisensory integration in complete

587   unawareness: evidence from audiovisual congruency priming. *Psychol. Sci.* **25**, 2006–

588   2016 (2014).

589   37. Salomon, R. *et al.* Unconscious integration of multisensory bodily inputs in the

590   peripersonal space shapes bodily self-consciousness. *Cognition* **166**, 174–183 (2017).

591   38. Stein, B. E. & Stanford, T. R. Multisensory integration: current issues from the

592   perspective of the single neuron. *Nat. Rev. Neurosci.* **9**, 255–266 (2008).

593   39. Serino, A., Farnè, A., Rinaldesi, M. L., Haggard, P. & Làdavas, E. Can vision of the body

594   ameliorate impaired somatosensory function? *Neuropsychologia* **45**, 1101–1107 (2007).

595   40. Dworkin, R. H. *et al.* Interpreting the clinical importance of treatment outcomes in

596   chronic pain clinical trials: IMMPACT recommendations. *J. Pain Off. J. Am. Pain Soc.* **9**,

597   105–121 (2008).

598

599

600

601

602   **Table and figure legends**

603
604
605   **Table 1: Patients clinical characteristics**
606

Gender	Age	Diagnostic	Months since SCS implant	Pulse width [us]	Amplitude [volt]	Frequency [Hz]	Illumination area (VR)	Pain laterality	Oswestry index
male	49	FBSS	9	180	7.60	40	both legs	both	31
female	45	FBSS	7	390	3.40	40	half leg	left	15
female	33	FBSS	4	300	3.30	60	1 leg	left	15
male	51	FBSS	4	320	4.60	70	half leg	right	21
male	40	FBSS	4	360	3.00	50	half leg	right	25
male	37	CRPS	8	300	2.00	95	1 leg	left	34
female	35	CRPS	10	450	4.10	80	1 leg	right	25
male	55	FBSS	3	130	6.30	40	both legs	both	20
male	52	CRPS	5	90	1.85	1000*	1 leg	right	31
male	61	FBSS	4	90	1.40	1000*	both feet	both	27
male	58	FBSS	5	90	3.20	1000*	both legs	both	20
male	34	CRPS	3	300	2.80	60	1 leg	left	11
male	55	CRPS	3.5	450	2.10	40	1 leg	left	5

female	53	FBSS	6	550	2.80	60	1 leg	left	17
female	57	FBSS	4	90	2.30	1000*	both legs	both	19

607   ***SCS high frequency (paresthesia-free) pattern**
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634   **Table 2: Inclusion and exclusion criteria**

635

636   Inclusion Criteria:

637   1. Age 18-and older at the time of enrollment
638   2. Patients carrying a diagnosis of CPRS or chronic refractory neuropathic leg pain
639   following FBSS
640   3. Patients who have implanted epidural SCS
641   4. The SCS implantation for at least three months prior to enrollment
642   5. Patients willing and able to provide informed consent

643   Exclusion Criteria:

644   1. Patients who are unable to effectively or efficiently communicate for example patients
645   suffering from speech deficits (dysarthria, aphasia) or are non-English speaking.
646   2. Patients with history of prior cranial surgery, significant brain lesions for example
647   intracranial tumors, strokes etc.
648   3. Evidence of untreated psychiatric disorders or drugs/alcohol abuse.
649   4. History of seizures

What it feels like to move via an intracortical brain machine interface (BMI)

Andrea Serino*[1,2,], Marcie Bockbrader*[3], Sam Colachis[4], Marco Solca[2], Tommaso Bertoni[1], Collin Dunlap[4], Kaitie Eipel[4], Patrick Ganze[4], Nick Annetta[4], Gaurav Sharma[4], Pavo Orepic[2], David Friedenberg[4], Per Sederberg[5], Nathan Faivre[2,6], Ali Rezai**[7], Olaf Blanke**[2,8]

1.MySpace Lab, Department of Clinical Neuroscience, University Hospital Lausanne (CHUV), Lausanne, Switzerland; 2. Laboratory of Cognitive Neuroscience, Brain Mind Institute & Center for Neuroprosthetics, Ecole Polytechnique Fèdèrale de Lausanne (EPFL), Campus Biotech, Geneva, Switzerland; 3.Department of Physical Medicine and Rehabilitation, The Ohio State University, Columbus, Ohio, US: 4 Medical Devices and Neuromodulation, Battelle Memorial Institute, Columbus, Ohio, US; Department of Psychology, University of Virginia, Charlottesville, Virginia, US. 6. Laboratoire de Psychologie et Neurocognition, Universitè Grenoble Alpes, Grenoble, France. 7. Rockefeller Neuroscience Institute, West Virginia University, Morgantown, West Virginia, US. 8. Department of Neurology, University Hospital, Geneva, Switzerland.

* Equal first author contribution; ** equal last author contribution

Abstract

Intracortical brain machine interfaces (BMI) decode motor commands from neural signals and translate them into actions, enabling movement for paralyzed individuals. The subjective sense of agency associated to BMI-generated actions, the involved neural mechanisms and its clinical relevance for BMI proficiency are currently unknown. By experimentally manipulating the coherence between decoded motor commands and sensory feedback in a tetraplegic BMI user, we demonstrate that primary motor cortex (M1) activity encodes sensory feedback, sensorimotor conflicts and subjective states of BMI actions. Neural signals processing the sense of agency affected the proficiency of the BMI system, underlining the clinical potential of the present approach. These new findings show that M1 encodes information related to action and sensing, but also sensorimotor and subjective agency signals, which in turn are relevant for BMI applications.

Main

When performing a voluntary movement, motor commands from the brain activate body effectors, which produce a cascade of reafferent sensory (proprioceptive, tactile, visual) cues. Motor commands are also associated with prediction signals about the sensory consequences of the movement. The congruency between motor commands, reafferent sensory feedback, and sensory predictions is at the basis of the sense of agency, our feeling of being in control of our actions (1–3). In case of damage to the motor system, motor commands that would trigger actions do not reach body effectors, leading to different types of paralysis, depending on the location and severity of damage. Intracortical brain machine interfaces (BMI) bypass such brain-body disconnection by decoding brain signals from different regions (i.e., primary motor cortex (M1), parietal or premotor cortex) and translating them into motor commands for the control of robots, exoskeletons (4, 5), neuromuscular functional electrical stimulation (6, 7) or other devices (8), enabling different actions (BMI actions) for patients with severe neuromotor impairments (9).

Despite major advances in intracortical BMIs based on research in human and non-human primates, the sense of agency for BMI actions, its neural mechanisms, and its impact on BMI performance is currently unknown. How does it feel to generate movements with a BMI – i.e., what is the sense of agency for BMI actions? Do motor neurons in human M1 encode not only motor commands, but also sensory feedback and agency signals for BMI actions? And does agency affect the efficiency of the BMI system - i.e. is agency of therapeutic benefit? We combined approaches from psychophysics, neurophysiology, neuroengineering and virtual reality (VR) and had the rare opportunity to investigate the above questions in a patient suffering from tetraplegia (caused by severe cervical spinal cord injury; C5/C6), who had been a BMI expert for two years before the start of the present study (6).

The BMI consisted of a 96-channel array implanted in the hand area of left M1 and actuated a transcutaneous forearm NMES system to translate decoded cortical signals

into right forearm and hand movements. To study the sense of agency and how it depends on sensory feedback, to gauge insight into the neural correlates of agency, and to evaluate its clinical impact, we experimentally manipulated the congruency between the decoded actions and the actions actuated by the BMI system. For this we manipulated the sensory feedback of the upper limb actions generated by the BMI-NMES (abbreviated as BMI actions) and carried out a series of experiments of 45 hours over a total of 13 days. As illustrated in Figure 1, the participant was instructed to realize a cued action with the BMI and was provided with movement-related sensory feedback using visual (via VR) and/or somatosensory (via NMES) stimulation. Critically, this feedback was either congruent or incongruent with respect to the motor commands decoded from M1: half of the trials, in which the decoded action corresponded to the cued action (e.g., open hand), were associated with congruent feedback (e.g., open hand), while the other half were associated with incongruent feedback (e.g. the opposite action: close hand). For each BMI action, we asked the participant whether he felt in control of that action and to rate his confidence about this judgement, allowing us to (1) gauge the sense of agency for BMI actions and how this was modulated by the congruency between motor commands and sensory feedback. Next, neural data from the M1 implant were analyzed to measure how (2) the sense of agency and (3) sensory feedback were encoded in the activity of M1 neurons, quantified as multi-unit (MU) firing rates and local field potentials (LFP). Finally, we assessed (4) the relevance of agency and sensory feedback for BMI performance. Investigating what it feels like to control actions mediated by an intracortical BMI, these data reveal for the first time the neural activity patterns in M1 (MU and LFP) that encode agency for BMI actions as well as the congruency and type of sensory feedback. We also show that the nature of somatosensory feedback and the related sense of agency affected the efficiency of the BMI system to correctly decode motor commands from M1, underlining the clinical relevance of sensory feedback and agency for future BMIs.

During the experiment, the participant was cued to execute one of four target actions (hand opening, hand closing, thumb extension, thumb flexion) using a validated BMI neuroprothesis. Neural activity corresponding to each target movement was recorded via

a 96-channel microelectrode array in M1 and a nonlinear support vector machine classifier was applied to decode the participant's chosen action from MU activity (see (*6*) for full description). On each trial, the classifier provided the likelihood of each target action (on a -1 to +1 range, in 100 ms bins), thus decoding one of the four target actions from the participant's M1 activity. In three different experiments, visual, somatosensory, or visual-somatosensory feedback about the BMI action was provided (Figure 1). In Experiment 1, VR was used to provide visual feedback, consisting of a life-size virtual arm on a monitor superimposed over the participant's right arm, matching the location and dimensions of the participant's real arm, which was occluded from view. In Experiment 2, NMES was used to provide 'somatosensory' feedback: the patient's upper limb muscles were electrically stimulated so he could feel, but not see the selected movement. Experiment 3 combined VR and NMES to provide 'visual-somatosensory feedback' (see below). In half of the trials, sensory feedback was congruent with the cued action, while in the other half it was incongruent (i.e., the opposite, action was executed) (see Figure 1B). At the end of each trial, we gauged the participant's sense of agency (0 or 1; Q1) and confidence (rating between 0 and 100; Q2). Importantly, the amount of sensory information was kept constant across experiments, by providing sham sensorimotor feedback in Experiment 1 (i.e., a pattern of NMES triggering no BMI action) and sham visual feedback in Experiment 2 (i.e., a static visual hand performing no action).
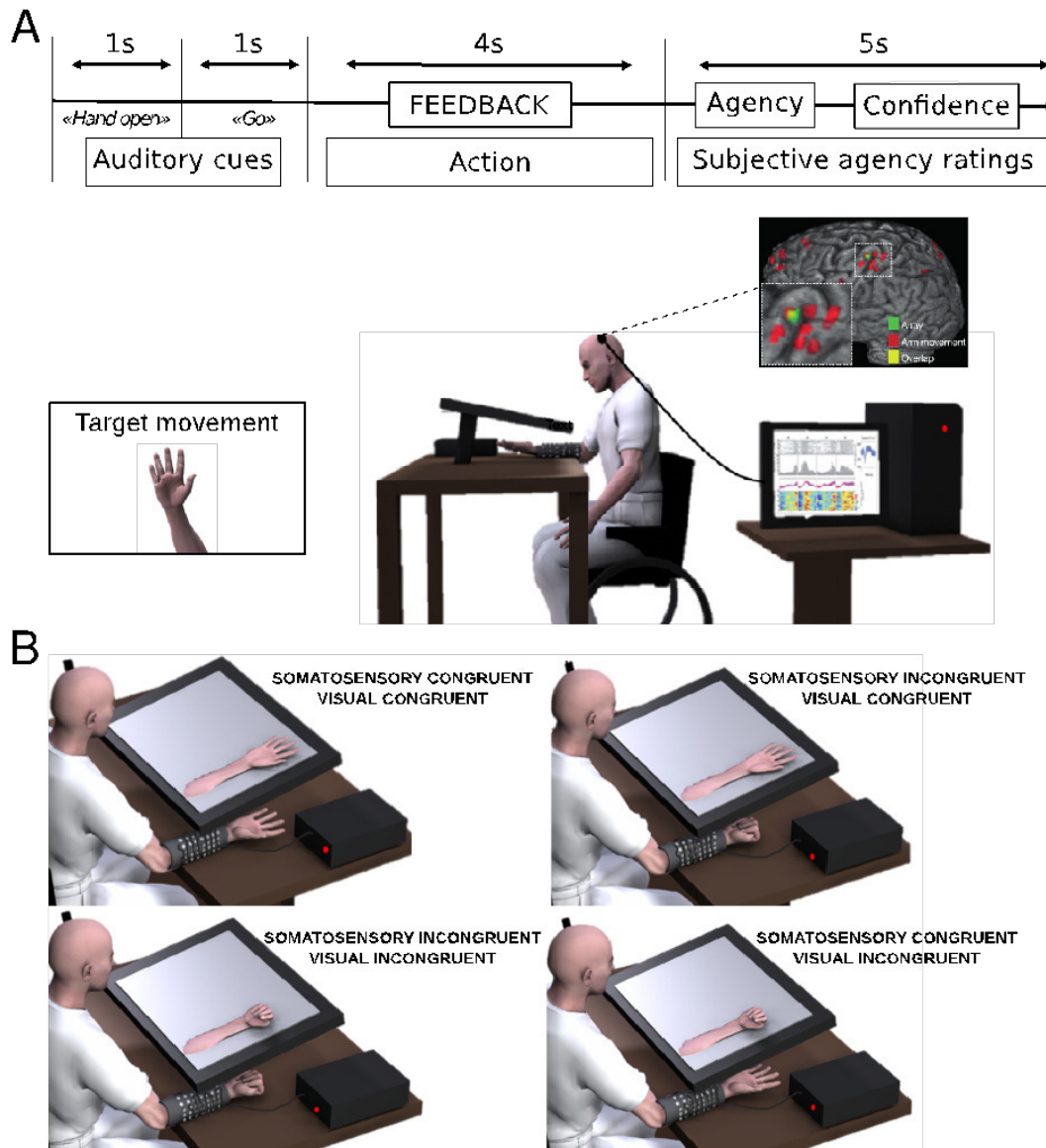
Figure 1. **Experimental setup.** A. Events during trials. One (out of four possible movements) was cued, following a "Go" signal to initiate the movement. The BMI classifier decoded the movement from M1 activity and sensory feedback was given. The patient answered two questions: Q1. "Are you the one who generated the movement?", by saying "Yes" or "No"; and Q2. "How confident are you?", by indicating a number ranging from 0 (absolutely unsure) to 100 (absolutely sure). B: Example of sensory feedback for one type of movement. The chosen movement was realized as a visual feedback, via virtual reality (VR – Experiment 1), as a somatosensory feedback, via NEMS (Experiment 2) or both (Experiment 3). In different congruency conditions, either the cued and correctly decoded movement (Congruent) or the opposite movement (Incongruent) was realized for the different modalities.

Results

Sensory feedback determines agency and confidence. Agency ratings were collected in a total of 844 trials (155, 243 and 448 trials for Experiments 1, 2 and 3, respectively; for Experiment 3 see below and supplementary material) and compared across feedback conditions using permutation tests. A null distribution of the mean agency rating was created by shuffling the condition labels over 10'000 iterations. P-values (2-sided) were estimated by counting the proportion of shuffled samples exceeding the observed average difference across conditions. As expected, and as shown in Figure 2A, we were able to manipulate agency and confidence for BMI actions. Thus, congruent visual (Experiment 1, 94% and 5.6% of positive responses to Q1 for congruent and incongruent trials, respectively, $p<.0001$) and congruent somatosensory (Experiment 2, 97.4% and 9% of positive responses for congruent and incongruent trials respectively, $p<.0001$) feedback resulted in more frequent agency responses versus incongruent conditions. Analyzing the role of feedback for confidence ratings (irrespective of the agency ratings), we found that confidence was not modulated by visual congruency (Experiment 1, mean Q2 rating = 70.9 for congruent, 73.6 for incongruent trials; $p = 0.28$), but by somatosensory congruency (Experiment 2, Q ratings were higher for somatosensory congruent [M = 74.1] than incongruent [M = 65] feedback; $p < 0.001$).

In order to disentangle the role of visual and somatosensory cues for agency and confidence, Experiment 3 combined VR and NMES including combinations of congruent and incongruent visual and somatosensory feedback (Figure 1). Most relevant are the comparisons between feedback conditions in which visual (V) and somatosensory (S) signals were both congruent or both incongruent (V-congruent/S-congruent; V-incongruent/S-incongruent) or when feedback was congruent in one modality and incongruent in the other modality (V-congruent/S-incongruent; V-incongruent/S-congruent). Results revealed that somatosensory congruency was more effective in driving the sense of agency and the associated confidence: ratings were stronger not only when both feedback signals were congruent (Q1 = 100% "Yes", mean Q2 = 83.8) as compared to both being incongruent (Q1 = 7.8% "Yes", mean Q2 = 72.3) (both p-value <

0.001), but also in the V-incongruent/S-congruent (Q1 = 68.5% "Yes", mean Q2 = 59.4) as compared to the V-congruent/S-incongruent condition (Q1 = 52.5% "Yes", mean Q2 = 54.8; p = 0.0035 and p = 0.036, for agency and confidence respectively) (Figure 2). Collectively, these data from Experiments 1-3 show that the congruency between decoded actions and sensory feedback, especially for the somatosensory modality, alters the sense of agency and confidence for actions mediated by an intracortical BMI.
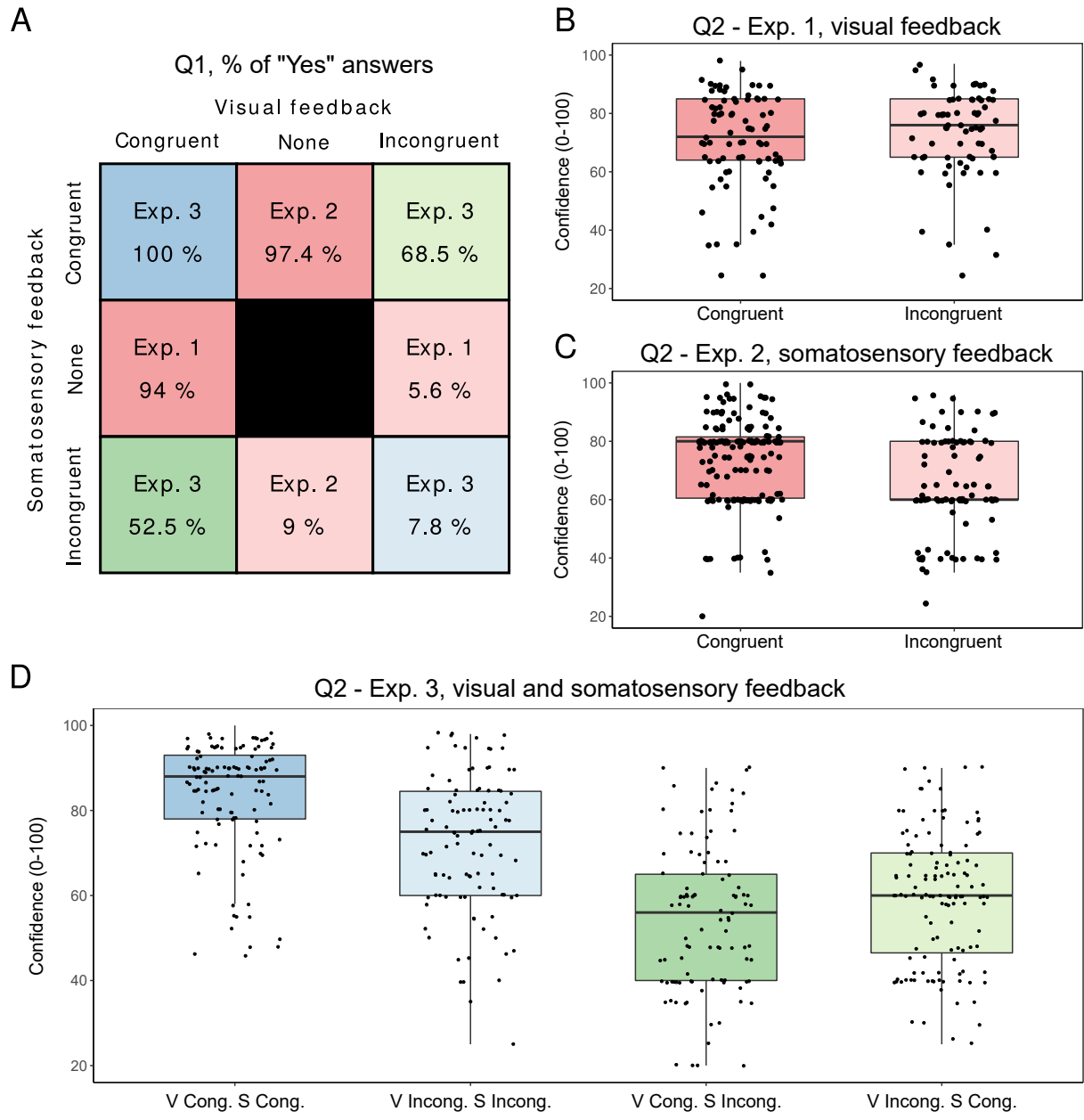
Figure 2. **Agency judgements and confidence depends on sensory feedback.** A. Proportion of "Yes" and "No" answers (Q1) to congruent and incongruent trials for the visual (Experiment 1) somatosensory (Experiment 2) and the combination of the two modalities (Experiment 3). B-C-D: Confidence about agency judgments. Distribution of Q2 responses as a function of the congruency of visual (B), somatosensory (C) or both (D) sensory feedback.

The sense of agency has been traditionally studied by presenting participants with different visuo-motor couplings ((*2*, *10–13*). In comparison, the role of somatosensory signals remains poorly understood (*14*), notably because it is normally impossible to decouple motor commands, somatosensory feedback and visual feedback, with extremely rare exceptions as in deafferented patients. Here we were able to contrast feedback cues that were congruent in one modality (e.g., visual) and incongruent in the other modality (e.g., somatosensory; and vice versa) with respect to the motor command and demonstrate that somatosensory cues dominate the sense of agency and the associated confidence for BMI-NMES actions. Of note, this effect cannot be due to the presence of somatosensory cues alone, as BMI actions in the visual condition were always associated with sham NMES stimulation producing somatosensory sensations without generating any actions (i.e., pseudo random somatosensory feedback, see supplementary material). Collectively these psychophysical data in a BMI expert reveal that agency for BMI actions depends on visual and somatosensory feedback (tactile and proprioceptive input) with somatosensory cues being more relevant.

Cortical signatures of sensory feedback in M1. We next investigated how such sensory feedback, that modulated the sense of agency, was encoded in M1 activity. We first analyzed the LFP amplitude in the different feedback conditions across the three experiments, using a regularized generalized linear model (ridge regression) and input signals from each individual channel at every time point (see Supplementary information). As shown in Figure 3A (left), the analysis distinguished congruent vs. incongruent visual feedback (maximum Cohen's Kappa K=0.42; p-value for the difference from baseline, <0.001) within a single period of a positive potential that lasted from ~700-1200 ms after the BMI action classification onset (Experiment 1). We could also distinguish congruent

vs. incongruent somatosensory feedback (maximum Cohen's Kappa K = 0.58; p < 0.0001) during two time periods: an early period characterized by a negative potential (stronger for incongruent feedback), starting at ~200 ms after BMI classification onset, followed by a later persistent differentiation lasting almost until the end of the trial. These results were further corroborated by data from Experiment 3: congruent trials in both modalities were clearly distinguished from incongruent trials in both modalities, lasting from ~250-1900 ms after BMI classification onset (maximum K = 0.66). In addition, V-congruent/S-incongruent trials were different from V-congruent/S-incongruent trials from ~300-1400 ms from BMI classification onset (maximum K = 0.31) (Figure 3B left). These findings show that visual and somatosensory feedback were both encoded by LFPs in human M1 and that such M1-LFP coding started earlier and was more stable over time for somatosensory feedback.

Applying the same decoding algorithm as for LFPs, we next determined if sensory feedback was also encoded by the spiking rate of MU in M1 (for methods see Supplementary material). As shown in Figure 3A (right), in Experiment 1, MU activity distinguished between congruent and incongruent visual feedback from ~400-900 ms from BMI classification onset (max K value = 0.41, p < .001). Extending LFP findings, an earlier and more stable differentiation between congruent and incongruent somatosensory feedback was found in MU activity in Experiment 2, with an effect as early as ~200 ms from the BMI classification onset (max K value = 0.66, p = < .001) and then persisted from 800 to 2000 ms. Similar results were found in Experiment 3 (Figure 3B, right), where MU activity distinguished between trials congruent and incongruent in both modalities and between V-congruent/S-incongruent and V-incongruent/S-congruent trials from ~160 ms from BMI classification onset. These data show that LFP and MU activity reflects visual and somatosensory feedback during actions driven by a BMI neuroprosthesis, with M1 activity reflecting somatosensory feedback starting ~200 ms after NMES activation (~150 ms after BMI classification onset, ~200 ms before M1 activity encoding visual feedback) and persisting for a longer period.
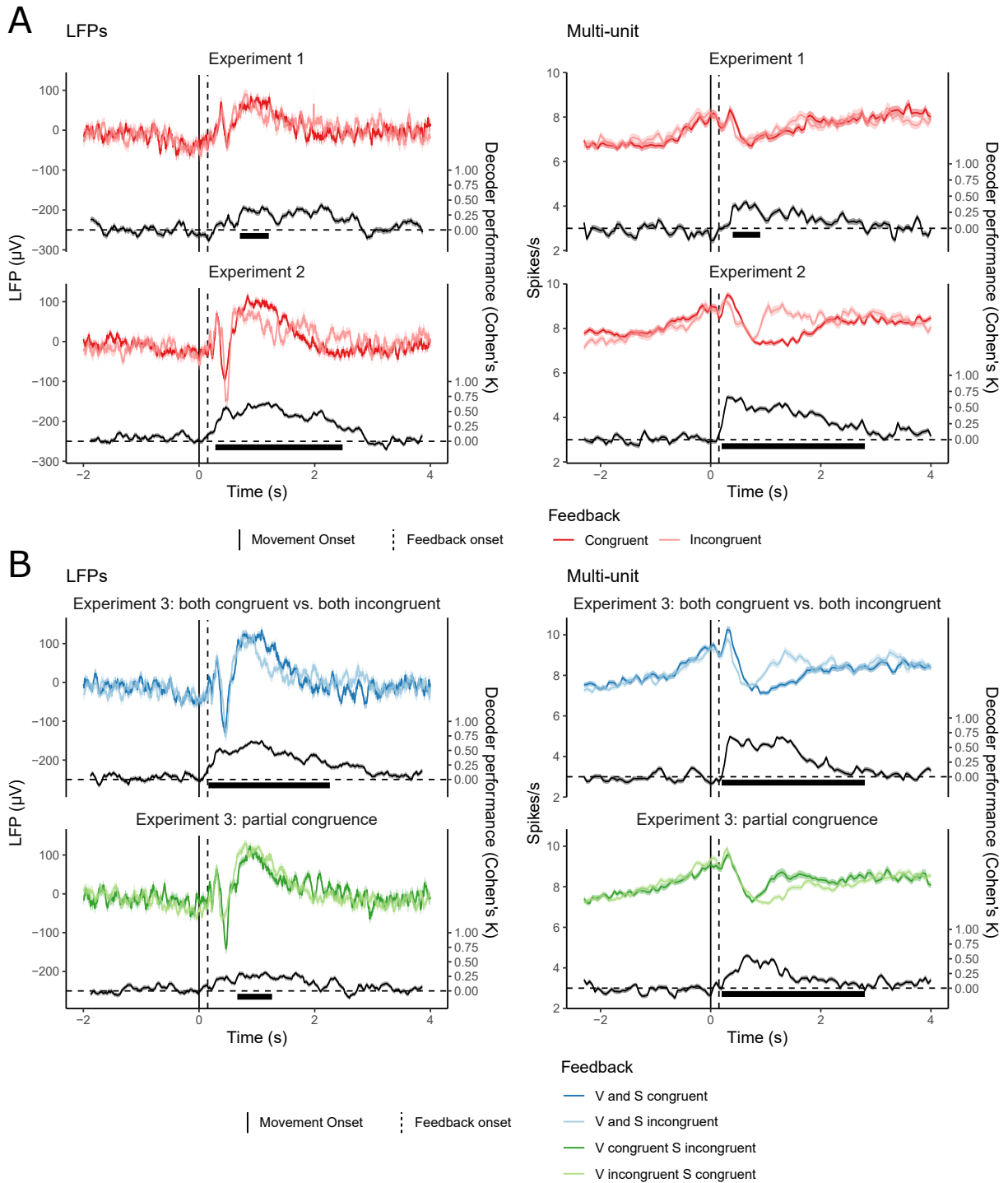
Figure 3. M1 activity depends on sensory feedback. Sensory feedback as encoded by Local field potentials (LFP; left panel) and Mult-iunit firing rates (MU; right panel). LFP and MU modulation for congruent and incongruent visual (Experiment 1) and somatosensory (Experiment 2) feedback (A) and for the combination of the two (Experiment 3, B). Colored lines represent averaged signal across all channels (shaded areas indicate SEMs); black

lines report the time-related k-values of the multivariate decoder distinguishing between congruent and incongruent feedback; the underlying thick segments indicate k-values significantly higher than chance level from cluster-based permutation analyses.

The role of somatosensory and visual information is an important topic in motor control, with robust evidence showing how perturbations of sensory feedback impact motor execution and adaptation (*15*). The present data show that the congruency between an intended action and somatosensory/visual feedback is encoded by M1 neurons at different latencies. To our knowledge, comparable data are not available in human or non-human primates, although previous studies in non-human primates described responses in M1 related to tactile and visual input (*16*, *17*), during active and passive movements (*18*) and during visual feedback of a pre-recorded movement (*19*, *20*). The present results are consistent with proposals that suggest that M1 activity codes both for movement types and their sensory consequences, in line with recent proposals describing how M1 neurons encode different movement parameters (see (*17*, *21*, *22*) for reviews). Here we report that, at the population level, human M1 activity in addition discriminates between arm movements that were congruent or incongruent with the motor command, as defined by somatosensory and visual feedback, with higher accuracy, earlier and more consistent processing for the former type of sensory information. Thus, neural coding in M1 contains, at the population level, information not only about the movement itself, but also about sensory consequences of actions, involving somatosensory-motor and visuo-motor loops.

Cortical signatures of the sense of agency in M1. Sensory-motor congruency is a key mechanism of agency for able-bodied actions; here we have shown that this also applies to agency and confidence for BMI-mediated actions and that LFPs and MU activity in human M1 distinguishes congruent vs. incongruent BMI actions. Next, we investigated to what extent LFP and MU activity in M1 also discriminate actions with and without an accompanying sense of agency. For each trial, we sorted LFP responses as a function of whether the participant reported agency or not. As seen in Figure 4A (left), LFP activity starting ~270 ms after BMI classification onset was found to code for agency and reached a maximum information value (K>.4) at ~1000 ms after BMI movement onset. Thus, BMI

actions for which the participant felt to be the agent were characterized by a different LPF pattern compared to BMI actions for which he did not. This was corroborated by MU activity analysis (Figure 4A, right). The MU firing rate was higher for trials with versus no agency; this discrimination started at ~300ms after BMI classification onset, until 500 ms, and peaked at ~400 ms (K max=.45). Later on, MU activity also differentiated for agency, with higher firing rate for trials with no agency (800-1600 ms after BMI classification). The same decoding was also able to discriminate trials with high vs. low confidence, based on a median split of Q2, from LFPs (max K = 0.296 at ~1200 ms) and MU (max K = 0.225 at ~400 ms).
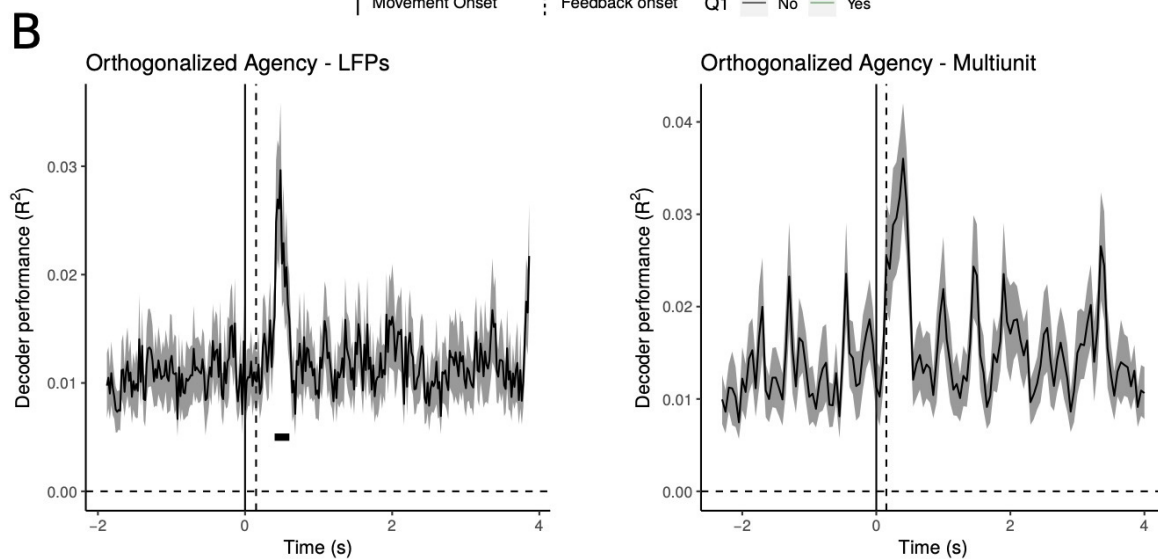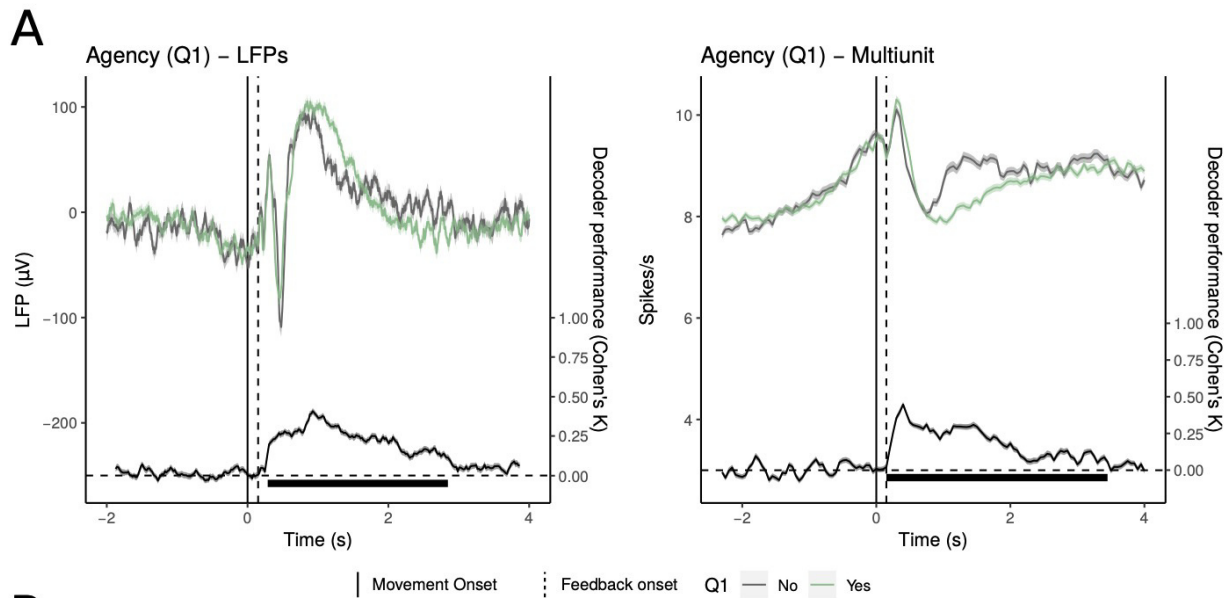
Figure 4. Sense of agency in M1. Sense of agency as coded by LFP (left) and Multi-unit firing rates (right). A. Left and right panels respectively show averaged LFP and Multi-unit modulation for high (green) and no (grey) agency response to Q1 (shaded areas indicate SEMs); black lines report the time-related k-values of the multivariate decoder distinguishing the two conditions; the underlying thick segments indicate k-values significantly higher than chance level from cluster-based permutation analyses. B: Results of the decoder discriminating between high vs. low orthogonalized agency scores from LFP (left) and MU (right) after regressing out for the effects of the congruency of sensory feedback and type of movements.

In the experimental design, sensory feedback congruency was used to modulate the sense of agency and this may have influenced these agency findings. Accordingly, we next tested whether LFP and MU contained information related to the sense of agency per se, after controlling for the effect of sensory feedback. For this we built a continuous measure of sense of agency and confidence allowing us to regress out the effect of sensory feedback. This new index was computed by recoding confidence ratings (Q2) as -Q2, for trails with no agency (as indicated in Q1) and +Q2 for trials with agency (from Q1). This index was then orthogonalized with respect to congruency in order to regress out this effect from the agency scores. As M1 signals also varied as a function of the different cued actions (see SI), the index was also orthogonalized for the type of action. We then used the same decoder to predict orthogonalized agency scores from LFP and MU activity over time. This analysis shows that LFPs predicted the sense of agency starting at ~450 ms after BMI classification onset ($p < 0.02$ with respect to baseline) (see Figure 4B left). A similar pattern was found when considering MU activity, although the peak failed to reach significance after cluster-based correction for multiple comparisons (Figure 4B right). These data show that M1 activity encodes the sense of agency and associated confidence level and was modulated by the congruency between motor commands and sensory feedback. Thus, subjective mental states associated with BMI actions and control are encoded by M1 activity at the LFP level (and to a minor extent at MU), independent of the neural processing associated with sensory feedback (see supplementary material for single channel analyses).

Somatosensory feedback modulates BMI classifier accuracy. Given the strong role of sensory congruency in determining agency and its coding in M1, we finally asked whether sensory feedback has any impact on the BMI classifier. To this aim, we tested whether the congruency between the decoded motor commands and sensory feedback (visual, somatosensory) affected the accuracy of the BMI classifier, defined as the summed suprathreshold activation values across a 4s window. In Experiments 1 and 2 we found that congruent somatosensory feedback improved classifier accuracy (t = 9.92; p < 0.0001) (Figure 5B right), while there was no effect due to visual feedback (p = 0.14) (Figure 5A). Moreover, incongruent somatosensory feedback was associated with lower classifier accuracy for the cued movement (Figure 5B left), and even increased classifier accuracy for the opposite movement (Figure 5B left). Thus, only somatosensory feedback congruency affected BMI accuracy in the present participant. This was extended by the results of Experiment 3, where we found a significant main effect of sensory feedback condition (F(3,444 = 15.83; p < 0.00001; Figure 5C). Further post-hoc corrected tests showed that the BMI classifier's accuracy was higher when feedback was congruent, than incongruent, in either modality (p < 0.0001). More interestingly, when feedback was congruent for the somatosensory modality and incongruent for the visual modality (V-incongruent/S-congruent) BMI accuracy was higher than in the opposite feedback condition (S-incongruent/V-congruent) (p < 0.001). These data from Experiments 1-3 show that BMI performance is affected by the congruency between the decoded motor commands and the somatosensory feedback induced by the action actuated by NMES. This finding is also coherent with the more reliable (i.e., earlier, more long-lasting and better decoded) processing of somatosensory feedback from M1 activity (LFP, MU). The fact that the same action as actuated by NMES (e.g., open hand) increased or decreased the BMI classifier performance, depending on whether somatosensory feedback was congruent (open hand) or incongruent (close hand) with the cued action, excludes that this effect was a generic artifact of NMES stimulation affecting the input to the BMI classifier independently from sensory information. Moreover, the finding that visual feedback did not alter BMI classifier accuracy shows that congruency per se cannot account for changes in BMI performance.

No prior study in humans and only few studies in monkeys directly tested the effects of sensory feedback on BMI performance (*19, 23*). Here we show, for the first time, an effect of feedback congruency on BMI performance, and the underlying role of M1 in this process. Our findings indicate that the recorded M1 units processed motor signals for the trained BMI actions, for sensory and sensory-motor signals reflecting the type and congruency of the sensory feedback. The improvement of BMI efficiency based on somatosensory feedback conditions is an important finding, considering that original BMI approaches for severely motor-impaired patients generally provide visual feedback only (*5, 24*) or somatosensory feedback by directly stimulating primary somatosensory cortex (*25–27*)(see *26 for a review*). Although from a single tetraplegic participant, the present data show that non-invasive somatosensory feedback via NMES not only enables higher subjective feeling of being in control (agency and confidence), but also leads to better actual control of the patient's BMI actions.
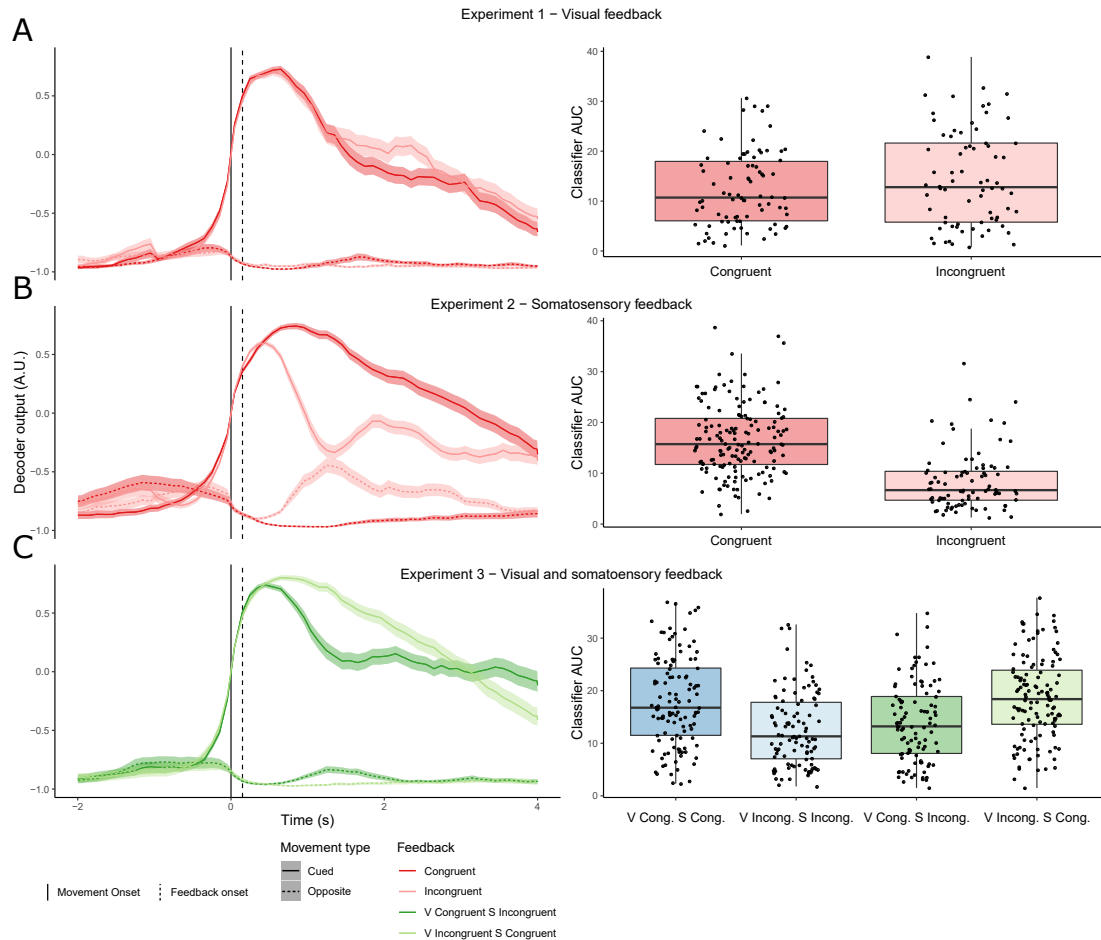

15

Figure 5. Performance of BMI classifier. The performance of the BMI classifier is modulated by the sensory feedback in the somatosensory modality. The left panels show the modulation in time of the performance of the BMI classifier for the 4 types of movements indicated as the cued movement (filled line) and the opposite (dashed line), as a function of feedback. The right panels show the area under the curve taken as index of global performance of the BMI classifier. The performance of the BMI classifier does vary not as a function of visual feedback (Experiment 1, A), whereas it is significantly better when somatosensory feedback is congruent both in Experiment 2 (B) and Experiment 3 (C).

Agency covaries with BMI classifier. We finally investigated whether agency has an impact on BMI efficiency and thus tested whether the sense of agency covaried with BMI classifier accuracy. We found that trials with agency versus trials without agency were associated with higher classifier accuracy. However, this was only the case when somatosensory (Experiment 2; $F(1,239=4.23$; $p<.05$), but not visual feedback was modulated (Experiment 1; $p=.14$), as also confirmed from analysis of data from Experiment 3 ($F(1,441=6.94$; $p<.001$). In addition, there was a significant correlation across all three experiments between BMI classifier accuracy and confidence (Q2, $F=46.95$; $p<.001$; $r^2=.10$). Thus, agency and confidence were both directly related to the performance of the present BMI system, but only when somatosensory feedback was involved. In order to confirm the role of agency on BMI performance, while controlling for other potential factors, we modelled classifier performance based on a multiple regression including agency, confidence, feedback type, feedback congruency, and movement type as regressors (Table 1 and see supplemental information). Our results show that for all the three experiments confidence covaried significantly with the performance of the classifier ($p<.01$; $<.01$; and $<.05$, respectively), even when the variability explained by the other factors was taken into account. As expected from the previous analyses, the congruency of the somatosensory ($p<.001$), but not of the visual ($p=.47$), feedback predicted the classifier's accuracy. Classifier performance also varied as a function of movement type (all p-values$<.01$). These findings show that movements with higher sense of agency and confidence are associated with higher BMI proficiency, suggesting that

subjective feelings associated to the control of a BMI-based neuroprosthesis is an important element to take into account to improve their effectiveness.

Predictors	Experiment 1			Experiment 2			Experiment 3		
	Estimates	CI	p	Estimates	CI	p	Estimates	CI	p
Intercept	0.96	-4.81 – 6.73	0.744	1.05	-3.07 – 5.17	0.618	6.75	1.84 – 11.66	**0.007**
Confidence	9.84	3.36 – 16.32	**0.003**	7.77	2.77 – 12.78	**0.003**	5.59	1.16 – 10.03	**0.014**
Q1=Yes	1.09	-3.23 – 5.42	0.621	1.24	-2.42 – 4.91	0.508	2.42	0.50 – 4.33	**0.014**
Visual Congruent	-1.58	-5.91 – 2.75	0.476						
Somatosensory Congruent				6.53	2.83 – 10.22	**0.001**			
Visual Congruent / Somatosensory Incongruent							-1.45	-4.02 – 1.12	0.271
Visual Incongruent / Somatosensory Congruent							2.81	0.55 – 5.06	**0.015**
Visual Incongruent / Somatosensory Incongruent							-2.89	-5.62 – -0.15	**0.039**
Cue HC	1.89	-0.95 – 4.72	0.194	1.42	-0.87 – 3.70	0.226	2.81	0.68 – 4.95	**0.010**
Cue TE	12.92	10.20 – 15.64	**<0.001**	3.39	1.13 – 5.66	**0.004**	6.20	4.25 – 8.16	**<0.001**
Cue TF	1.32	-1.89 – 4.54	0.421	6.33	3.98 – 8.67	**<0.001**	7.41	5.43 – 9.38	**<0.001**
Trial number	0.03	-0.08 – 0.13	0.617	-0.05	-0.13 – 0.03	0.204	-0.02	-0.09 – 0.05	0.600
Observations	155			243			448		
R^2 / adjusted	0.500 / 0.476			0.394 / 0.376			0.241 / 0.225		

Table 1. **Sense of agency.** Multiple regression coefficients predicting agency scores, while regressing out the effects of sensory feedback and movement type.

Discussion

By combining techniques from neurophysiology, neuroengineering, and VR with psychophysics of agency, we were able to study for the first time the sense of agency for actions enabled by a BMI-based neuroprosthesis and found that congruent sensory feedback boosted agency and confidence when controlling BMI actions. Moreover, we showed that human M1 processes not only motor and sensory information, but also different levels of congruency between sensory and motor signals and the resulting sense of agency. The present data are also of clinical relevance, because our NMES-based BMI approach, by providing congruent somatosensory feedback (without direct S1 stimulation)

to a tetraplegic patient, improved the ability of the BMI classifier in decoding the patient's motor commands. Interestingly, such higher BMI proficiency was associated with a stronger sense of agency, suggesting that, beyond supporting close-loop systems and M1 feedback in general, somatosensory feedback and signals related to subjective aspects of motor control (i.e. agency) are important input for improving BMI proficiency. Quantifying subjective action-related mental states and including controlled motor and sensory feedback may therefore provide new levels of comfort and personalization and should be considered for the design of future BMIs.

The present data demonstrate that M1 activity contains information specifically linked to subjective aspects of motor control, in particular the sense of agency and confidence that our participant associated with his BMI actions. It is known that agency likely involves a network of multiple brain areas from which we did not record in the present study (e.g., posterior parietal cortex (*28*) and angular gyrus; anterior insula (*29, 30*); supplementary motor cortex (*31*); premotor cortex (32) for review see (*3, 33*)). However, our findings – even if coming from a single tetraplegic patient - directly demonstrate that M1 activity contains sufficient information to decode actions for which a human participant feels to be in control. This finding is important not just for the field of neuroprosthetics and its clinical goals, but also for basic neuroscience as well as current ethical and legal debates about the subjective sense of agency and responsibility when applying neurotechnology solutions for human repair or enhancement (*33*-35).

References

1. S. J. Blakemore, D. M. Wolpert, C. D. Frith, Abnormalities in the awareness of action. *Trends Cogn. Sci.* **6**, 237–242 (2002).
2. M. Jeannerod, *Motor cognition : what actions tell the self* (Oxford University Press, 2006; https://books.google.fr/books?hl=en&lr=&id=__IJnFPtwQoC&oi=fnd&pg=PR5&dq=marc+j eannerod&ots=iM6BMThNmQ&sig=75vEbwO8uAhphkY2LSu0pNe-3jw#v=onepage&q=marc jeannerod&f=false).
3. P. Haggard, Sense of agency in the human brain. *Nat. Rev. Neurosci.* **18**, 196–207 (2017).
4. L. R. Hochberg, D. Bacher, B. Jarosiewicz, N. Y. Masse, J. D. Simeral, J. Vogel, S. Haddadin, J. Liu, S. S. Cash, P. van der Smagt, J. P. Donoghue, Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*. **485**, 372–375 (2012).

5.	J. L. Collinger, B. Wodlinger, J. E. Downey, W. Wang, E. C. Tyler-Kabara, D. J. Weber, A. J. McMorland, M. Velliste, M. L. Boninger, A. B. Schwartz, High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet*. **381**, 557–564 (2013).

6.	C. E. Bouton, A. Shaikhouni, N. V. Annetta, M. A. Bockbrader, D. A. Friedenberg, D. M. Nielson, G. Sharma, P. B. Sederberg, B. C. Glenn, W. J. Mysiw, A. G. Morgan, M. Deogaonkar, A. R. Rezai, Restoring cortical control of functional movement in a human with quadriplegia. *Nature*. **533**, 247–250 (2016).

7.	A. B. Ajiboye, F. R. Willett, D. R. Young, W. D. Memberg, B. A. Murphy, J. P. Miller, B. L. Walter, J. A. Sweet, H. A. Hoyen, M. W. Keith, P. H. Peckham, J. D. Simeral, J. P. Donoghue, L. R. Hochberg, R. F. Kirsch, Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration. *Lancet*. **389**, 1821–1830 (2017).

8.	M. A. Lebedev, M. A. L. Nicolelis, Brain-Machine Interfaces: From Basic Science to Neuroprostheses and Neurorehabilitation. *Physiol. Rev.* **97**, 767–837 (2017).

9.	J. P. Donoghue, Connecting cortex to machines: recent advances in brain interfaces. *Nat. Neurosci.* **5**, 1085–1088 (2002).

10.	G. Knoblich, N. Sebanz, Agency in the face of error. *Trends Cogn. Sci.* **9**, 259–261 (2005).

11.	A. Sato, A. Yasuda, Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*. **94**, 241–255 (2005).

12.	S. Marchesotti, R. Martuzzi, A. Schurger, M. L. Blefari, J. R. del Millán, H. Bleuler, O. Blanke, Cortical and subcortical mechanisms of brain-machine interfaces. *Hum. Brain Mapp.* **38**, 2971–2989 (2017).

13.	N. Evans, S. Gale, A. Schurger, O. Blanke, Visual Feedback Dominates the Sense of Agency for Brain-Machine Actions. *PLoS One*. **10**, e0130019 (2015).

14.	M. Tsakiris, M. R. Longo, P. Haggard, Having a body versus moving your body: Neural signatures of agency and body-ownership. *Neuropsychologia*. **48**, 2740–2749 (2010).

15.	S. H. Scott, T. Cluff, C. R. Lowrey, T. Takei, Feedback control during voluntary motor actions. *Curr. Opin. Neurobiol.* **33**, 85–94 (2015).

16.	S. Shokur, J. E. O'Doherty, J. A. Winans, H. Bleuler, M. A. Lebedev, M. A. L. Nicolelis, Expanding the primate body schema in sensorimotor cortex by virtual touches of an avatar. *Proc. Natl. Acad. Sci.* **110**, 15121–15126 (2013).

17.	N. G. Hatsopoulos, A. J. Suminski, Sensing with the Motor Cortex. *Neuron*. **72**, 477–487 (2011).

18.	N. G. Hatsopoulos, Q. Xu, Y. Amit, Encoding of movement fragments in the motor cortex. *J. Neurosci.* **27**, 5105–14 (2007).

19.	A. J. Suminski, D. C. Tkach, A. H. Fagg, N. G. Hatsopoulos, Incorporating Feedback from Multiple Sensory Modalities Enhances Brain-Machine Interface Control. *J. Neurosci.* **30**, 16777–16787 (2010).

20.	D. Tkach, J. Reimer, N. G. Hatsopoulos, Congruent activity during action and action observation in motor cortex. *J. Neurosci.* **27**, 13241–50 (2007).

21.	M. M. Churchland, K. V. Shenoy, Temporal Complexity and Heterogeneity of Single-Neuron Activity in Premotor and Motor Cortex. *J. Neurophysiol.* **97**, 4235–4257 (2007).

22. A. B. Schwartz, Movement: How the Brain Communicates with the World. *Cell*. **164**, 1122–1135 (2016).

23. J. E. O'Doherty, M. A. Lebedev, P. J. Ifft, K. Z. Zhuang, S. Shokur, H. Bleuler, M. A. L. Nicolelis, Active tactile exploration using a brain–machine–brain interface. *Nature*. **479**, 228–231 (2011).

24. L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, J. P. Donoghue, Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*. **442**, 164–171 (2006).

25. G. A. Tabot, J. F. Dammann, J. A. Berg, F. V Tenore, J. L. Boback, R. J. Vogelstein, S. J. Bensmaia, Restoring the sense of touch with a prosthetic hand through a brain interface. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18279–84 (2013).

26. S. N. Flesher, J. L. Collinger, S. T. Foldes, J. M. Weiss, J. E. Downey, E. C. Tyler-Kabara, S. J. Bensmaia, A. B. Schwartz, M. L. Boninger, R. A. Gaunt, *Sci. Transl. Med.*, in press, doi:10.1126/scitranslmed.aaf8083.

27. S. J. Bensmaia, L. E. Miller, Restoring sensorimotor function through intracortical interfaces: progress and looming challenges. *Nat. Rev. Neurosci.* **15**, 313–325 (2014).

28. M. Desmurget, K. T. Reilly, N. Richard, A. Szathmari, C. Mottolese, A. Sirigu, Movement Intention After Parietal Cortex Stimulation in Humans. *Science (80-.).* **324**, 811–813 (2009).

29. C. Farrer, C. D. Frith, Experiencing Oneself vs Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency. *Neuroimage*. **15**, 596–603 (2002).

30. V. Chambon, D. Wenke, S. M. Fleming, W. Prinz, P. Haggard, An Online Neural Substrate for Sense of Agency. *Cereb. Cortex*. **23**, 1031–1037 (2013).

31. I. Fried, R. Mukamel, G. Kreiman, Internally Generated Preactivation of Single Neurons in Human Medial Frontal Cortex Predicts Volition. *Neuron*. **69**, 548–562 (2011).

32. Fornia, L. *et al.* Direct electrical stimulation of the premotor cortex shuts down awareness of voluntary actions. *Nat. Commun.* **11**, 705 (2020).

33. M. Sperduti, P. Delaveau, P. Fossati, J. Nadel, Different brain structures related to self- and external-agency attribution: a brief review and meta-analysis. *Brain Struct. Funct.* **216**, 151–157 (2011).

34. R. Yuste, S. Goering, B. A. y Arcas, G. Bi, J. M. Carmena, A. Carter, J. J. Fins, P. Friesen, J. Gallant, J. E. Huggins, J. Illes, P. Kellmeyer, E. Klein, A. Marblestone, C. Mitchell, E. Parens, M. Pham, A. Rubel, N. Sadato, L. S. Sullivan, M. Teicher, D. Wasserman, A. Wexler, M. Whittaker, J. Wolpaw, Four ethical priorities for neurotechnologies and AI. *Nature*. **551**, 159–163 (2017).

35. I. Fried, P. Haggard, B. J. He, A. Schurger, Volition and Action in the Human Brain: Processes, Pathologies, and Reasons. *J. Neurosci.* **37**, 10842–10847 (2017).

36. Shanechi, M. M. Brain–machine interfaces from motor to mood. *Nat. Neurosci.* **22**, 1554–1564 (2019).

Methods

<u>Participant</u>

The participant in this study was enrolled in a pilot clinical trial (NCT01997125, Date: November 22, 2013) of a custom neural bridging system (Battelle Memorial Institute) to reanimate paralyzed upper limbs after C4-6 spinal cord injury. The system consisted of a Neuroport data acquisition system (Blackrock Micro, Salt Lake, Utah), custom signal processing and decoding algorithms (Battelle), and a NeuroLife Neuromuscular Stimulation System (Battelle). The trial received investigational device exemption (IDE) approval by the US Food and Drug Administration and Institutional Review Board approval through the Ohio State University (Columbus, Ohio). The study conformed to institutional research requirements for the conduct of human subjects. The site of the experiments was the Ohio State University NeuroRehabLab (Bockbrader, PI) and data was analyzed at Ohio State (Columbus, Ohio) and École polytechnique fédérale de Lausanne (EPFL, Switzerland). The participant provided informed consent at the time of enrollment and also provided written permission for photographs and videos.

The study participant was a 22 year-old male at the time of study enrollment. He had complete C5 ASIA A, non-spastic tetraparesis from cervical spinal cord injury associated with a diving accident 3 years prior. On neurological exam, he had full motor function bilaterally for C5 level muscles (e.g., biceps and shoulder girdle muscles), but no motor function below the C6 level. He had 1/5 strength on the right and 2/5 strength on the left for wrist extension (C6 level) on manual muscle testing. His sensory level was C6 on the left and C5 on the right, although he had sensation for pressure on his right thumb. He had preserved proprioception for shoulder, elbow and wrist joint position, but was at chance level for distinguishing digit joint positions (flexion/neutral/extension) for the thumb and fingers. He had mild finger flexor contractures bilaterally, limiting finger extension at the proximal and distal interphalangeal joints of digits 2-5.

He was implanted with a 4.4 x 4.2mm intracortical silicon Neuroport microelectrode array (Blackrock Microsystems) in the dominant hand/arm area of his motor cortex on 4/22/2014, as previously described (6). The implant site was determined by preoperative functional neuroimaging obtained while the participant visualized movements of his right

hand and forearm. He began using cortically-controlled transcutaneous neuromuscular electrical stimulation (NMES) on his right forearm on 5/23/14, participating in sessions to practice device use for up to 3.5 hrs/day and 3 days/week. In 7/2015, his practice with the device was reduced to 2 days/week. Data for this study was collected over 13 sessions (45 hours) from 11/16/2016 - 2/20/2017, corresponding to post-implant days 939-1035. One session with visual and NMES feedback was used for practice (5 blocks of 32 trials on post-implant day 939). At the time of data collection, the participant was an expert brain-machine interface (BMI) user with over 800 hours of study participation.

Of note, the participant underwent cognitive testing of attention, memory and processing ability (without the BMI) approximately one year after Utah array implantation (January – July, 2015). He scored in the gifted range with superior verbal abilities, attention, and working memory (ranging between 92nd - 99th percentile for his age), and no significant differences between auditory or visual memory. However, his processing speed and performance scores were significantly affected by his upper limb impairment (ranging between 27th – 39th percentile for his age).

Cortical Signal Acquisition And Classification

Neural data (96 channels) were acquired from the left motor cortex Utah array through the Neuroport data acquisition system (Blackrock Micro). Raw data were processed using analog hardware with 0.3Hz 1st order high-pass and 7.5kHz 3rd order Butterworth low-pass filters, then digitized at 30,000 Hz. Data were divided into 100ms bins and passed into Matlab (version 2014b), where signal artifact was removed by blanking over 3.5ms around artifacts (defined as signal amplitude >500µV at the same time on 4 of 12 randomly-selected channels). Signals were decomposed into mean wavelet power (MWP) using the 'db4' wavelet over 100ms (37). Coefficients within the multiunit frequency bands (234–3,750Hz, coeficients of scales 3, 4, 5, 6) were averaged across the 100ms window and normalized by channel (by subtracting the mean and dividing by the standard deviation of each channel and scale, respectively). Normallized coefficients for each channel were averaged across scales 3-6, creating 96 MWP values (one for each channel) per each 100ms. MWP values were fed as features into a real-time, nonlinear

support vector machine (SVM) classifier (38) with five classes (hand open, hand closed, thumb extension, thumb flexion, and rest). Classifier activation values were computed for each 100ms bin and ranged from -1 to 1. Classifier output represented the movement pattern (hand open, hand closed, thumb extension, thumb flexion) with the highest activation greater than threshold (zero). If no movement classes had activation greater than zero, the classifier was in the "rest" state. If multiple output classes exceeded threshold, only the one with the highest score was used to provide feedback.

Signal quality was stable (39) during the interval of data collection; but represented about a 30% decline in MWP normalized to post-implant 87 (40). (See below for single unit statistics.) Average impedance was approximately 200 kΩ, a decline of 40% of its initial value. Average signal-to-noise was approximately 17.5dB, a decrease of about 10% of its initial value (41). Most of the decline in signal quality occurred in the first 400 days post-implantation.

Classifier Training And Neurally Controlled Hand Movements

Before each session, the SVM classifier was trained in an adaptive manner over 5 blocks. Each block consisted of 3 repetitions of 4 movements (hand open, hand closed, thumb extension, thumb flexion) presented in a random order. Movements were cued for 3-4s (4-5s inter-cue interval) using a small, animated hand in the corner of the video display. Feedback was given with both NMES and the feedback hand on the video screen. During the first training block, scripted feedback was provided simultaneously with the cued movements. In subsequent blocks, appropriate movements were activated when an output class for a given movement exceeded threshold (>0). Training took approximately 10-15 minutes per session.

Neuromuscular Electrical Stimulation

The NMES system was used to evoke hand and finger movements by stimulating forearm muscles. The system consisted of a multi-channel stimulator and a flexible, 130-electrode, circumferential forearm cuff. Coated copper electrodes with hydrogel interfaces (Axelgaard, Fallbrook, CA) were 12mm in diameter, spaced at regular intervals in an array

(22mm longitudinally X 15mm transversely), and delivered current in monophasic, rectangular pulses at 50Hz (pulse width 500μs, amplitude 0-20mA). Desired hand/finger movements were calibrated at the beginning of each session by determining/confirming the intensity and pattern of electrodes required to stimulate intended movements. This took 5-10 minutes per session.

During the experiment, the participant's view of NMES-evoked movements was obscured from view by the video display. During Experiment 1, sham NMES feedback was given (current at an intensity equivalent to what was used for movement calibration patterns, but that did not evoke movement). During Experiments 2 and 3, NMES feedback was provided that evoked hand and finger movements.

Virtual Reality Animation

A physics-based animated hand was used to provide visual feedback of classifier activation. During training, two animated hands were displayed, a small cue hand at the bottom left and a larger centrally-placed feedback hand (Figure 1 main text). During the experiment, the display was oriented over the participant's forearm, a single, centrally-placed feedback hand was displayed at approximately the size and location of the participant's right hand (the cue hand was not displayed). During Experiments 1 and 3, feedback was provided using the virtual hand. During Experiment 2, sham visual feedback was given (the feedback hand remained in a neutral, rest position).

Feedback Congruency

In half of the trials across Experiments, the visual and/or somatosensory feedback was covertly manipulated to be incongruent with the cue. In incongruent trials, when the participant correctly activated the classifier associated with the cue, he received feedback opposite to the cue (i.e., hand closed for "hand open", thumb extension for "thumb flex", etc.). In congruent trials, he received feedback consistent with the cue (i.e., hand open for "hand open", thumb flexion for "thumb flex", etc.).

Agency Assessment

All experimental trials began with a verbal cue ("hand open", "hand closed", "thumb extend", "thumb flex"), followed by a 2 second delay, then a verbal cue ("go"). During the next 4s, the participant was given feedback based on classifier activation levels, and then was told to "stop"). Over the next 5-5.5s, the participant reported whether he felt in control of the movement ("yes" or "no") and his degree of certainty (0-100). The next trial began at the end of this 5-5.5s interval. There were 32 trials per block in Experiments 1 and 2 and 26 trials per block in Experiment 3.

Trial Selection and Time-locking

To ensure that the participant is succesfully activating the classifier for the cued movement, and the signal can be meaningfully time-locked to movement onset, we applied the following selection criteria on the trials. We consider it as a correct imagined movement when the participant is able to maintain the classifier of the cued movement above the threshold for at least 600 ms (6 classifier output bins). We retain trials in which at least one correct movement happens between the GO cue and 1.5 seconds before the STOP cue. Epochs are then constructed by time-locking every trial with respect to the onset of such imagined movements. In case several correct movements occurred during the same trial, the time-locking is relative to the first movement. Furthermore, we excluded 128 trials from the session on which the participant systematically reported problems with controlling the BMI system and absent subjective agency. Globally, we retained 846 out of 1408 trials (60%).

Note that, since we define the onset as the beginning of the 100 ms bin of neural activity that is fed to the classifier, and around 50 ms are required to compute the output, the corresponding feedback is received about 150 ms after the onset of the imagined movement.

Experiment 1: Agency Assessment with Virtual Hand Feedback and Sham NMES

Twelve blocks of 32 trials were collected on post-implant days 953 (4 blocks), 988 (4 blocks), and 1035 (4). In each trial, the participant received a verbal cue to perform a movement ("hand open", "hand closed", "thumb extend", "thumb flex"). When a classifier

crossed threshold during the 4 second feedback window, feedback was given by showing movement of the virtual hand and by activating sham NMES (radial wrist electrode activation that did not elicit movement, did not vary from trial to trial, and that the participant could feel and distinguish from real NMES feedback). Feedback on half of the trials was randomly selected to be incongruent with the cue. His subjective sense of agency and level of certainty were recorded for each trial.

A total of 384 trials were collected across three days. After removing trials where the participant did not respond correctly by activating the classifier associated with the cue, and the session on post-implant day 1035 (see trial selection paragraph), 83 congruent and 72 incongruent trials remained for behavioral and neural activity analysis.

Experiment 2: Agency Assessment with NMES Feedback and Sham Virtual Hand

Twelve blocks of 32 trials were collected on post-implant days 941 (5 blocks), 960 (3 blocks), and 967 (4 blocks). In each trial, the participant received a verbal cue to perform a movement ("hand open", "hand closed", "thumb extend", "thumb flex"). When a classifier crossed threshold during the 4 second feedback window, feedback was given by activating movement of the participant's hand and wrist through NMES and showing sham visual feedback (non-moving hand). The participant could not see his own hand/wrist, but could distinguish his hand state based what the stimulation patterns felt like to him. Feedback on half of the trials was randomly selected to be incongruent with the cue. His subjective sense of agency and level of certainty were recorded for each trial.

A total of 384 trials were collected across three days. After removing trials where the participant did not respond correctly by activating the classifier associated with the cue, 154 congruent and 89 incongruent trials remained for behavioral and neural activity analysis.

Experiment 3: Agency Assessment with Virtual Hand and NMES Feedback

Twenty blocks of 32 trials were collected on post-implant days 993 (3 blocks), 990 (5 blocks), 1007 (4 blocks), 1014 (3 blocks), and 1021 (5 blocks). In each trial, the participant received a verbal cue to perform a movement ("hand open", "hand closed", "thumb

extend", "thumb flex"). When a classifier crossed threshold during the 4 second feedback window, feedback was given by activating movement of the participant's hand and wrist through NMES and showing movement of the virtual hand. The participant could not see his own hand/wrist, but could distinguish his hand state based what the stimulation patterns felt like to him. Congruency with respect to the cue was manipulated independently in the visual and somatosensory modalities such that 25% of the trials were each: congruent for both visual and NMES feedback, incongruent for both visual and NMES feedback, congruent for visual but incongruent for NMES feedback, congruent for NMES but incongruent for visual feedback. His subjective sense of agency and level of certainty were recorded for each trial.

A total of 520 trials were collected across five days. After removing trials where the participant did not respond correctly by activating the classifier associated with the cue, the number of trials that remained for behavioral and neural activity analysis were: 117 congruent for both visual and NMES feedback, 103 incongruent for both visual and NMES feedback, 101 congruent for visual and incongruent for NMES feedback, and 127 congruent for NMES and incongruent for visual feedback.

Firing Rate Calculation and Single Unit Analyses

Single units were identified through offline data processing. For each block, raw voltage recordings at each channel were processed in a series of steps. First, FES stimulation artifact was removed using a 500µV threshold and 3.5ms artifact removal time window. The removed window was replaced with an interpolated segment to retain temporal information. Then, the raw signal with FES artifact removed was processed with a 300-3000Hz bandpass filter. The filtered data was fed into an automated spike detection and sorting algorithm, wave_clus (42) using the default optimization settings. A threshold was set to four times the standard deviation of the noise and used to detect spike locations. A wavelet decomposition was performed on the spikes to extract features and a superparamagnetic clustering algorithm was used to cluster the spikes into groups, representative of individual single units. The superparamagnetic clustering algorithm was used to eliminate spikes that were considered noise to ensure only single units were

analyzed. As spike sorting was not performed before data collection, there was no way to match single units across days. Additionally, the number of single units detected at a given channel fluctuated between days, possibly due to micro-movement of the array and brain state changes. For this reason, all single units detected at a given channel were considered the same, and pooled at the single channel level as multiunit activity in subsequent analysis.

Offline neural decoding

Sensory feedback congruency and subjective ratings (Q1 and Q2) were decoded offline both from LFPs and from multiunit activity. For LFP analysis, the signal amplitude for each channel was downsampled to 500 Hz, band-passed between 0.1 and 12 Hz with an IIR filter, and smoothed using sliding averaging windows of 250 ms. Following multiunit spike times calculation (see above), multiunit firing rate was estimated at 20 Hz over a 250 ms sliding window.

We fed each channel's signal amplitude (LFP) or firing rate (multiunit) as predictors to a penalized linear decoder based on ridge regressions (43). A separate model was trained to decode congruency (Q1) or confidence (Q2) on each signal timepoint, with a sampling rate of 20 and 500 Hz for multiunit and LFPs respectively. Decoding performance was evaluated by computing and averaging Cohen's k (logistic regression; Q1) or R^2 (linear regression; Q2) values over 10 independent 10-fold cross validation runs. The regression was performed through the "train" function of the R "caret" package (44). To evaluate the statistical significance of the decoding, we generated a null decoding performance distribution by applying the same decoding methods on the data after randomly shuffling Q1 and Q2 values. 1000 permutations were generated, and the decoding performance was evaluated for each of them. Then, a t-value was assigned to every time-point both in real and permuted data, by comparing its decoding performance to the null distribution of permuted data. Finally, the t-values were used to define significant decoding time windows based on a cluster-based permutation test on each epoch's largest cluster (45) . After checking that the t-value threshold used to define clusters was not significantly affecting the results, its value was set at 2.

Methods references

37. S. Mallat. *A Wavelet Tour of Signal Processing. A Wavelet Tour of Signal Processing*. (2009) https://doi.org/10.1016/B978-0-12-374370-1.X0001-8

38. C. Humber, K. Ito, C. Bouton. Nonsmooth Formulation of the Support Vector Machine for a Neural Decoding Problem. *ArXiv*. (2010) http://arxiv.org/abs/1012.0958 81

39. S. C. Colachis, M. A. Bockbrader, M. Zhang, D. A. Friedenberg, N. V. Annetta, M. A. Schwemmer, … G. Sharma. Dexterous control of seven functional hand movements using cortically-controlled transcutaneous muscle stimulation in a person with tetraplegia. *Frontiers in Neuroscience*, *12*(APR), 1–14. (2018) https://doi.org/10.3389/fnins.2018.00208 92

40. S. Colachis. Optimizing the Brain-Computer Interface for Spinal Cord Injury Rehabilitation. (2018)
http://rave.ohiolink.edu/etdc/view?acc_num=osu1524083218999623

41. M. Zhang, M. A. Schwemmer, J.cE. Ting, C. E. Majstorovic, D. A. Friedenberg, M. A. Bockbrader, … G. Sharma. Extracting wavelet based neural features from human intracortical recordings for neuroprosthetics applications. *Bioelectronic Medicine*, *4*(1), 11. (2018) https://doi.org/10.1186/s42234-018-0011-x

42. R. Q. Quiroga, Z. Nadasdy, Y. Ben-Shaul. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, *16*(8), 1661–1687. (2004) https://doi.org/10.1162/089976604774201631

43. A. E. Hoerl, R. W.Kennard. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. (1970) https://doi.org/10.1080 / 00401706.2000.10485983

44. M. Kuhn. R Package: caret, Ver. 6.0-80. *CRAN*. (2018)

45. E. Maris, R. Oostenveld. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. (2007) https://doi.org/10.1016/j.jneumeth.2007.03.024

Acknowledgments

Formal analysis, Methodology, Visualization, Review & editing; AR: Funding acquisition, Resources, Supervision, Review & editing; OB: Conceptualization, Funding acquisition, Methodology, Supervision, Writing.

Competing interests: AS, MS, TB, NF, PS, OB: no conflict of interest.CD, KE, PG, GS, NA, DF have patents for the BMI system.

Data and materials availability

Behavioral data and processed data necessary to reproduce the figures in the main text can be found in the OSF repository accessible at:

https://osf.io/7rma5/?view_only=9928bd8e32a748828f7ecfdbeb1f8baa

Supplementary Materials

Supplementary Text

Figs. S1 to S5

Supplementary Materials

Movement awareness in human M1: evidence from an intracortical Brain Machine Interface, virtual reality and Neuromuscular Electrical Stimulation

Andrea Serino*, Marcie Bockbrader*, Sam Colachis, Marco Solca, Tommaso Bertoni, Collin Dunlap, Kaitie Eipel, Patrick Ganzer, Nick Annetta, Gaurav Sharma, David Friedenberg, Per Sederberg, Nathan Faivre, Ali Rezai, Olaf Blanke

Correspondence to: andrea.serino@unil.ch

This PDF file includes:

Supplementary Text
Figs. S1 to S5

Supplementary Text

20 <u>Whole Sample Results</u>
To ensure that our trial selection criteria is not biasing the behavioral results on reported agency and confidence, we run the same analyses as in the main text on the whole sample of 1408 trials. In this sample, we still exclude 128 trials from session of post-implant day 1035 (see trial-selection paragraph). All results go in the same direction as in 25 the main text, therefore we succintly report them without further discussion.
In Experiment 1, Q1 and Q2 are significantly higher in the congruent condition $p < 0.001$ and $p=0.003$ respectively. In Experiment 2, Q1 and Q2 are significantly higher in the congruent condition $p < 0.001$ and $p<0.001$ respectively. In experiment 3, when contrasting Vis. congruent/NMES incongruent and Vis. incongruent/NMES congruent, we 30 found Q1 and Q2 to be higher in the Vis. incongruent/NMES congruent condition ($P=0.009$ and $p=0.02$ respectively).

<u>Sensory congruency and agency at the level of single channels.</u>
We analysed whether sensory feedback and agency was more specifically 35 processed in any of the 96 channels from the M1 implant. To this aim, we firstly identified the channels where the decoder's coefficients for MU activity more strongly and significantly contributed to the decoding of the of visual and somatosensory congruency (see Figure S1). Two of the 6 significant channels for visual and somatosensory feedback overlapped (channels 80 and 76 in figure S1). Sense of agency was more strongly 40 decoded from 7 channels, two of them overlapping with both visual and somatosensory congruency decoding (channels 80 and 76), and 3 others with somatosensory congruency decoding only, confirming the stronger interdependency between somatosensory signals of agency judgments. All these electrodes were mainly located in the rostral part of the array. Interestingly, the electrodes more strongly decoding sensory 45 congruency and agency were clearly dissociated from those more strongly decoding for the intended BMI action, since the spatial distributions of the decoder coefficients for the type of intended movements highlighted significant electrodes in the caudal part of array, not overlapping with sensory congruency nor agency electrodes (e.g., channels 17 and 67 in figure S1). Thus, M1 activity, also at the single channels level, codes not only for 50 type of movement, but also for the congruency between selected movement and sensory feedback, and the associated sense of agency. Despite stronger contribution from specific electrodes, additional analysis suggests that both sensory congruency and agency are more likely to be encoded at the population level as the power of the decoder in classifying congruent vs. incongruent movements or high vs. low agency actions was 55 higher at the population level than at any of the best 20 channels (see Figure S2).
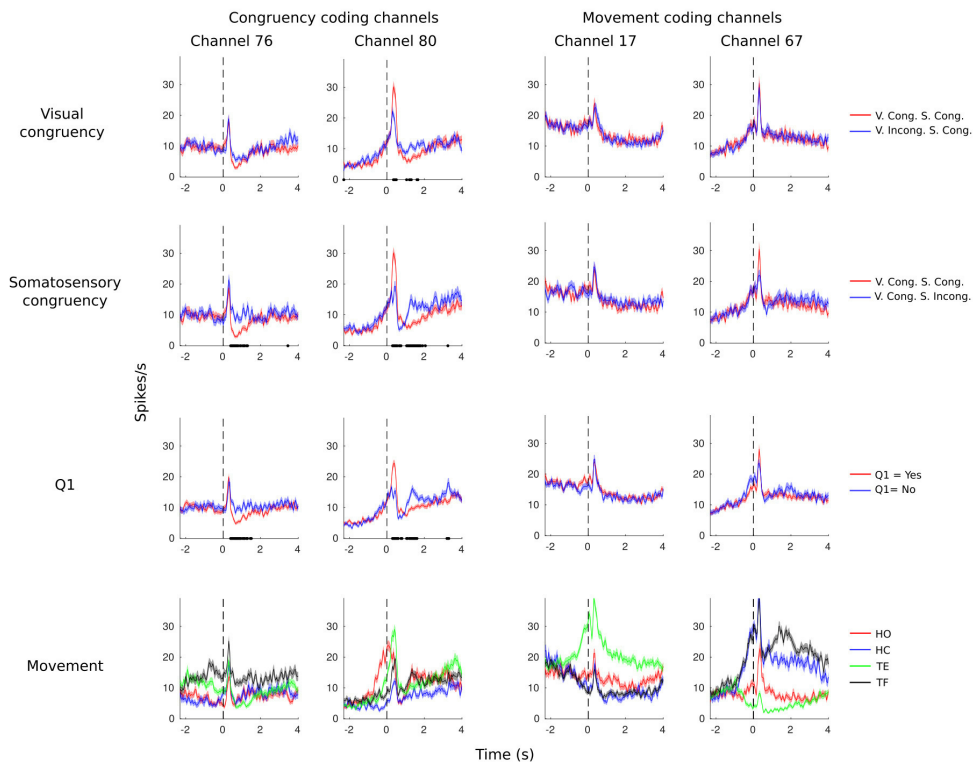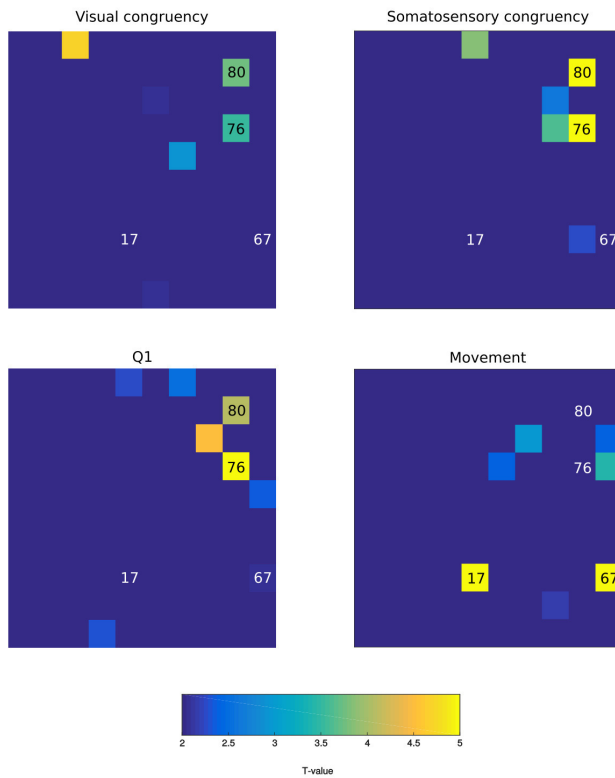
Fig. S1.

Contribution of individual channels to neural decoding. In the top panel we show a map of individual channels that contributed most to decoding visual congruency, somatosensory congruency, agency and movement type (going left to right and then top to bottom). To isolate the role of visual and somatosensory modalities in feedback congruency, while minimizing the signal variability between sessions, we trained another ridge regression based decoder on experiment 3, so that we can compare the two modalities in the same set of trials. For visual congruency, we trained the decoder on trials with congruent somatosensory feedback, by contrasting congruent and incongruent visual feedback. For visual congruency, we contrasted congruent and incongruent somatosensory feedback in trials with congruent visual feedback. For movement, to maintain a 2 class decoding schema, we contrasted extension movement (hand open, thumb extension) and flexion movements (hand close, thumb flexion), in trials with congruent visual and somatosensory feedback. To evaluate the contribution of each channel we compared its coefficient in the Ridge regression in a 1 second window starting at movement onset (where decoding of all features is significant), with the distribution of coefficients over all the 96 channels on a 1 second window preceding movement onset, used as a null distribution. T-values are extracted and thresholded at 2. Then, their absolute value is color-coded and displayed on the array grid. Note that this method aims at setting a cut-off on each channel's contribution to the neural decoding, for easier visualization, not at providing a statistically rigorous estimate of decoding significance. In the lower panel we show exemplary channels' response to different conditions. In the two columns on the left we show the response of "congruency coding" channels 76 and 80. In the two columns on the right the same is done for "movement coding" channels 17 and 67. Going from the top to the bottom row, we contrast visual and somatosensory congruencies, positive and negative agency ratings, and the four different movements. After movement onset, the two congruency coding channels clearly differentiate feedback congruency and agency, but show no big difference with respect to the movement. Conversely, two movement coding channels show large differences with respect to the movement even prior to movement onset, suggesting motor intention coding, but no modulation from feedback congruency or agency. Shaded areas indicate standard errors, and black dots indicate significant differences after FDR correction (only where two conditions are contrasted).
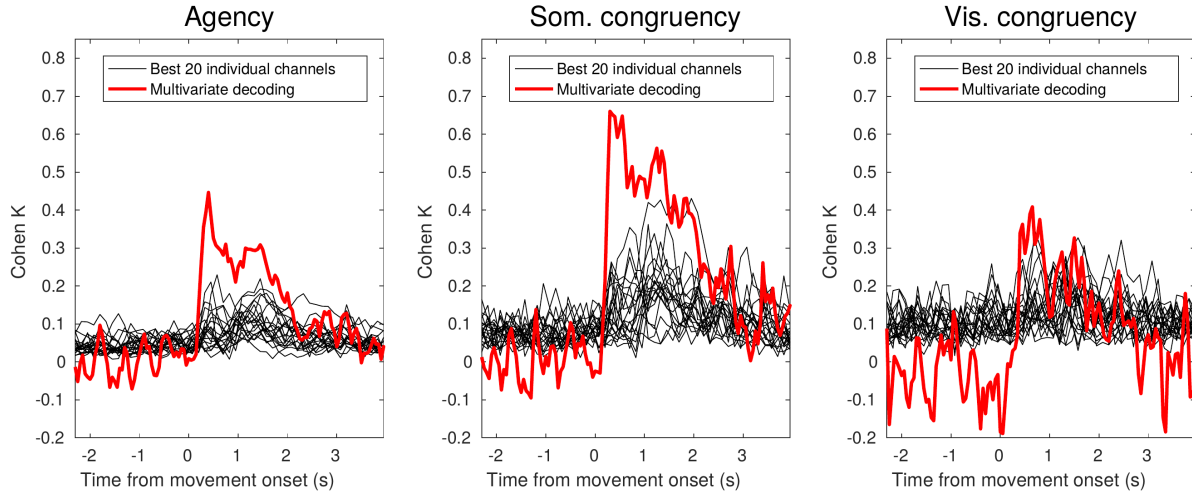
Fig. S2.

Comparison between multivariate and single-channel decoding, for Agency, somatosensory and visual congruency (from left to right). Red lines represent cross-validated Cohen's K values for the multivariate Ridge regression presented in the main text. Black lines represent Cohen's K of univaried decoding based on the 20 channels giving the highest mean K value. Note that, to be more conservative, the single channel decoding is not cross-validated, and therefore its performance is slightly overestimated. For the same reason, chance level is higher than 0 and pre-movement decoding is slightly above 0 in the univariate case. Nevertheless, multivariate decoding is greatly overperforming univariate decoding in the case of agency and somatosensory congruency, and only slightly better in the case of visual congruency.
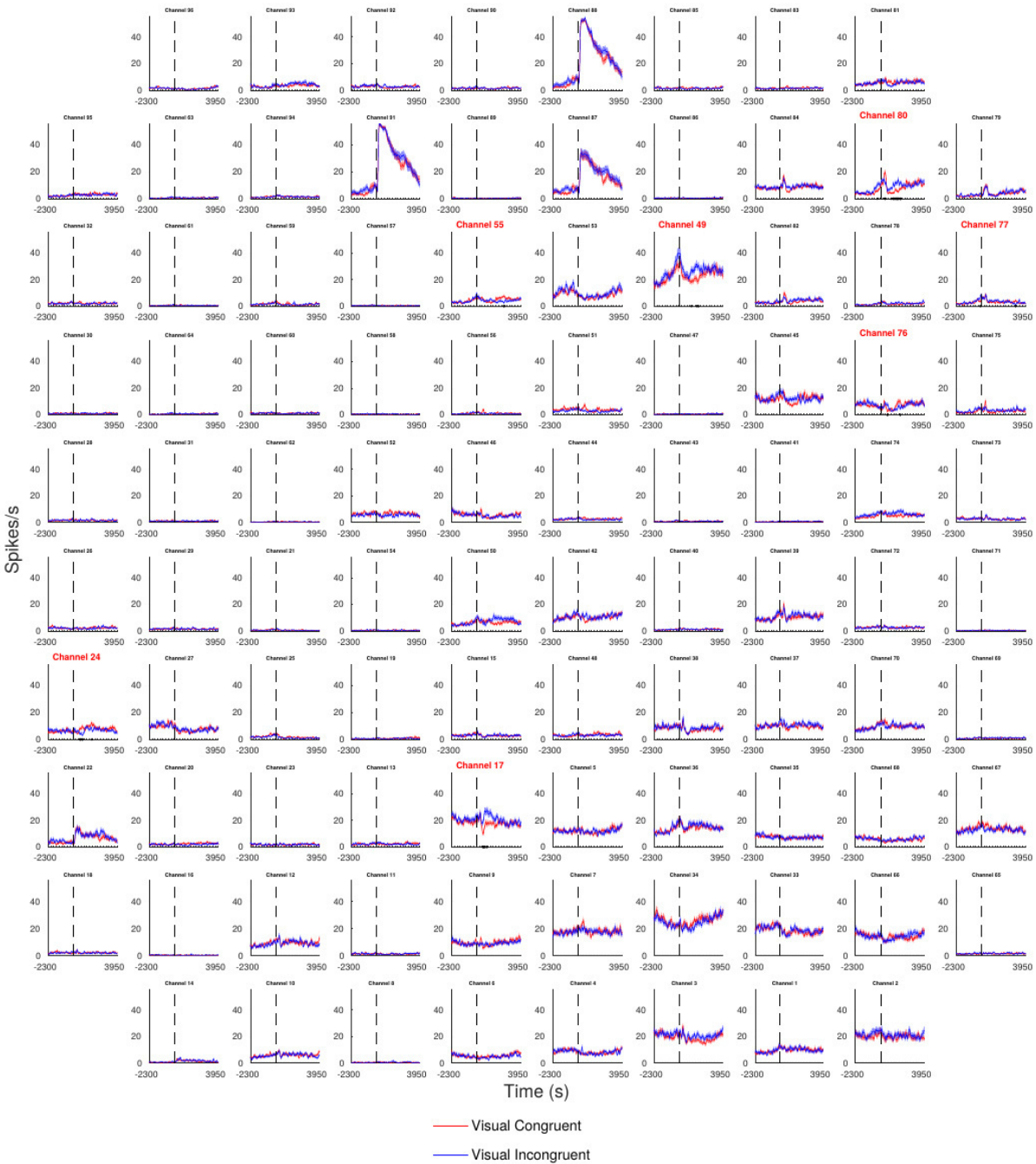
Fig. S3.

Time-locked multiunit response of individual channels contrasting visual congruent and visual incongruent feedback in Experiment 1. The black lines indicate significantly different responses between the two conditions (FDR corrected across timepoints), and subplots with red titles indicate channels with at least one significant timepoint.

6
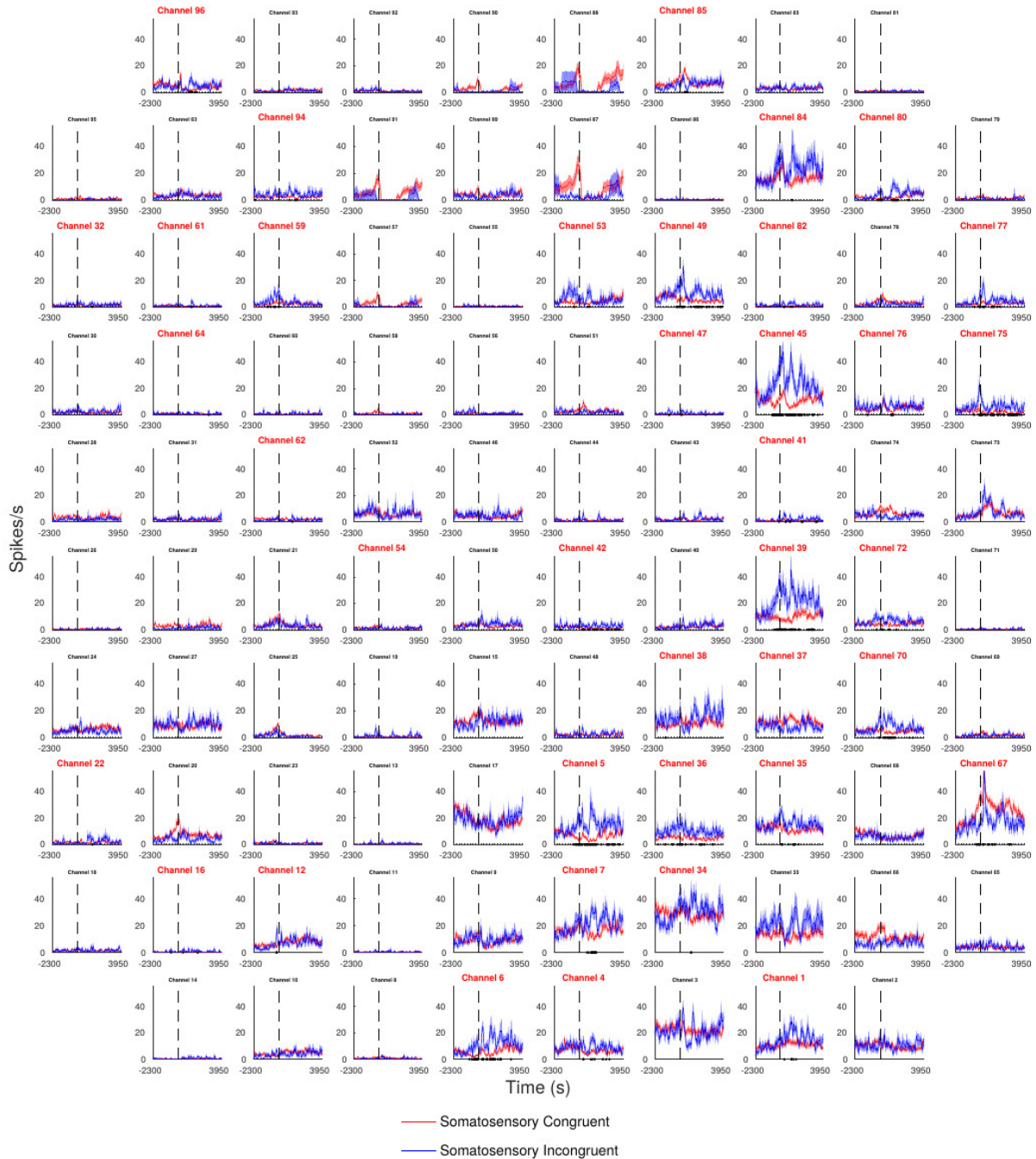
Fig. S4

Time-locked multiunit response of individual channels contrasting somatosensory congruent and somatosensory incongruent feedback in Experiment 2. The black lines indicate significantly different responses between the two conditions (FDR corrected across timepoints), and subplots with red titles indicate channels with at least one significant timepoint.
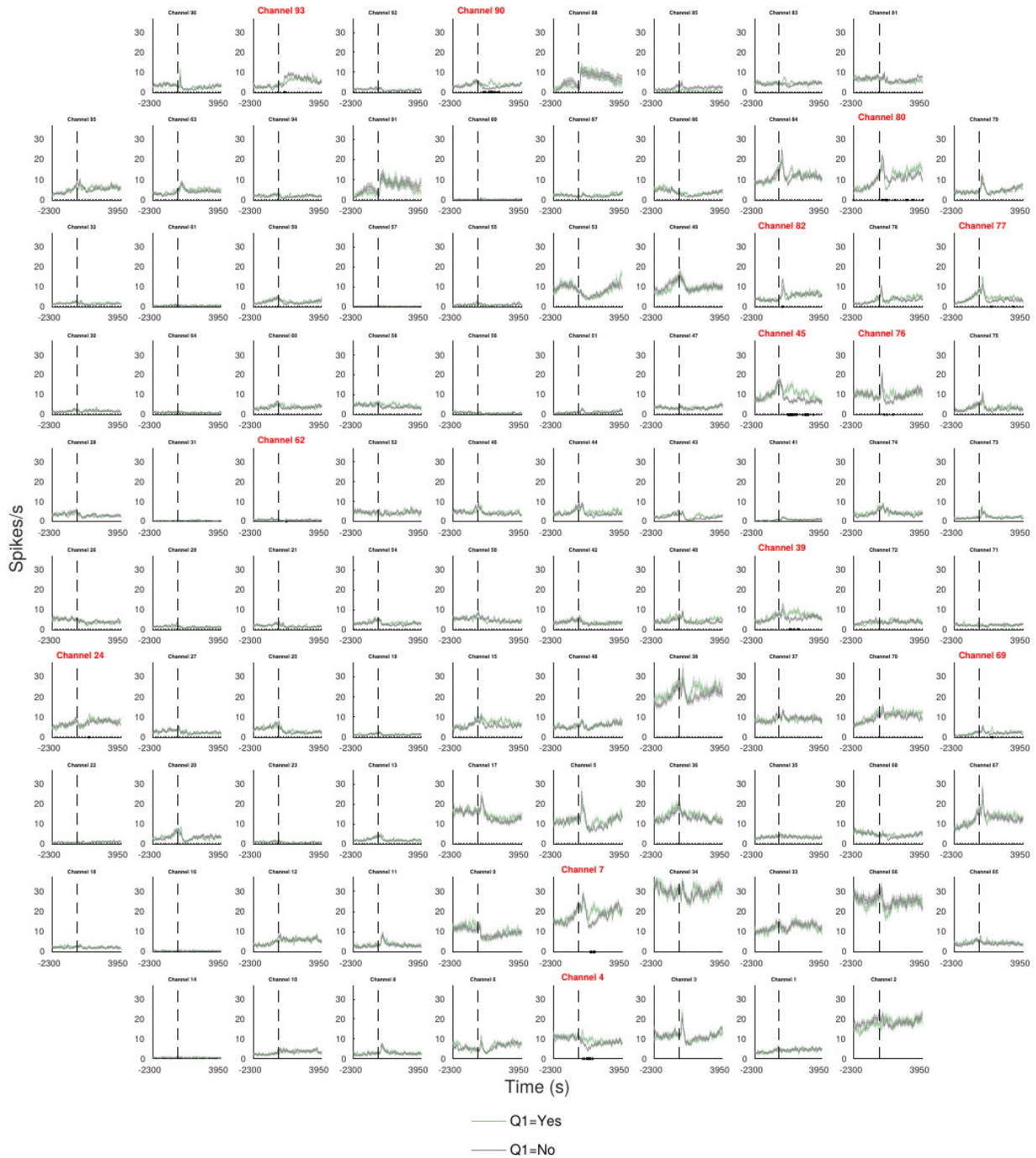
Fig. S5
Time-locked multiunit response of individual channels contrasting positive and negative sense of agency in all experiments. The black lines indicate significantly different responses between the two conditions (FDR corrected across timepoints), and subplots with red titles indicate channels with at least one significant timepoint.

Curriculum Vitae

Personal information

First name / Last name	**Pavo Orepić**
Home Address	Rue Dr César-Roux 9, 1005 Lausanne, Switzerland
E-mail	pavo.orepic@gmail.com
Nationality	Croatian
Date of birth	28th March 1993
Gender	Male
E-mail	pavo.orepic@gmail.com
Phone	+41 (0)76 721 17 44
OrcID	0000-0001-6965-7578

Education

November 2016 – December 2020	**PhD in Neuroscience**
Thesis	Dissecting self-voice perception: From bone conduction to robotically-induced self-other voice misattribution in healthy listeners
PhD advisor	Prof. Olaf Blanke
Date of the defense	December 11th 2020
Institution	Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland
September 2014 – September 2016	**Master of Science in Computer Science**
Thesis	Augmented virtuality through body detection and integration used in neuroscience research
Thesis advisor	Prof. Stephanie Speidel
Institution	Karlsruhe Institute of Technology (KIT), Germany
September 2011 – July 2014	**Bachelor of Science in Computer Science**
Thesis	Data encryption with ultra-low power consumption in wireless transmission
Thesis advisor	Assist. Prof. Martin Žagar
Institution	Faculty of Electrical Engineering and Computing (FER), University of Zagreb, Croatia

Work experience

November 2016 – December 2020	**PhD student**
Advisor	Prof. Olaf Blanke
Institution	Laboratory of Cognitive Neuroscience, Swiss Federal Institute of Technology Lausanne, Switzerland
Main activities	Designing and conducting behavioral and neuroimaging (EEG) experiments Analysis of behavioral, physiological (respiration, ECG, EDA) and EEG data Teaching assistance at university courses (total 312 hours) and supervision of student projects
November 2016 – May 2017	**Visiting scholar**
Advisor	Dr. Ali Rezai
Institution	Center for Neuromodulation, Department of Neurosurgery, The Ohio State University (OSU), USA
Main activity	Development of RealiSM, a virtual reality (VR) platform used in neurosurgery research
October 2015 – April 2016	**Research assistant**
Advisor	Prof. Rüdiger Dillmann
Institution	Humanoids and Intelligence Systems Lab, Karlsruhe Institute of Technology (KIT), Germany

Main activity	Development of surgical assistance systems
June 2015 – March 2016	**Research assistant**
Advisor	Prof. Wolfgang Rosenstiel
Institution	The FZI Research Center for Information Technology, Karlsruhe Institute of Technology, Germany
Main activity	Optimization of automotive electronics

Supervision and mentoring

February 2020 – present	**Master Thesis Project** of Dorian F. A. Garin
March 2019 – April 2019	**Master "Lab immersion" Project** of Arthur Barakat
July 2019 – August 2019	**Bachelor Thesis Project** of Melissa Faggella

Teaching activities

February 2019 – June 2019	**Neuroscience III**, Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland Preparation and supervision of exercise classes and examination
February 2019 – June 2019 February 2018 – June 2018	**Genomics and Bioinformatics**, Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland Preparation and supervision of exercise classes and examination
September 2017 – January 2018	**Sensorimotor Neuroprosthetics**, Swiss Federal Institute of Technology Lausanne, Switzerland Preparation and supervision of exercise classes and examination
March 2014– July 2014	**Computer Aided Design of Electronic Systems**, University of Zagreb, Croatia Laboratory assistant
September 2011 – February 2012	**Digital Logic,** University of Zagreb, Croatia Laboratory assistant

Conference talks

September 2019	**"Sensorimotor conflicts louden quiet voices in healthy listeners"** Early Career Hallucination Research (ECHR), Durham, The United Kingdom

Poster presentations

February 2020	**"How a robot can make you hear voices: Sensorimotor conflicts alter self-voice perception in healthy listeners"** The Swiss Society for Neuroscience (SSN), Bern, Switzerland
June 2019	**"Sensorimotor conflicts related to somatic passivity experiences louden quiet voices in healthy listeners"** The Association for the Scientific Study of Consciousness (ASSC), London, Ontario, Canada
February 2019	**"Sensorimotor conflicts related to somatic passivity experiences louden quiet voices in healthy listeners"** The Swiss Society for Neuroscience (SSN), Geneva, Switzerland

Memberships in scientific societies

February 2019 – present	Swiss Society for Neuroscience (SSN)
March 2018 - present	National Center of Competence in Research (NCCR) Synapsy

Awards and fellowships

March 2020	Finals of EPFL's "My Thesis in 180 seconds" competition, Audience award
September 2019	Early Career Hallucination Research (ECHR) Travel Bursary, Fellowship

June 2019	Swiss Society for Neuroscience (SSN) Travel Fellowship, Fellowship
March 2018	Best oral presentation at the Doctoral Students in Life Sciences (ADSV) retreat, Audience award
April 2016	Swiss-European Mobility Programme, Scholarship
September 2014	ERASMUS scholarship, Scholarship
September 2012, September 2013	University of Zagreb Excellence Scholarship for top 10% talented students, Scholarship

Outreach activities

March 2020	**"How a robot can make you hear voices: Sensorimotor conflicts alter self-voce perception in healthy listeners"** My thesis in 180 seconds (MT180) competition, Lausanne, Switzerland (video)
March 2018	**"Voices and Robots"** Association of Doctoral Students in Life Sciences (ADSV) retreat, Ovronnaz, Switzerland
April 2017	**"Neuroscientific Perspectives on Apparitions, Out-of-body experiences and Self-healing"** Invited talk, Parapsychological Association, Columbus, Ohio, USA (video)

Extracurricular Activities

January 2018 – January 2020	**President of the Social Campus Biotech Association (SCBA)**
Institution	Campus Biotech, Geneva, Switzerland
Main activity	Managing social activities at Campus Biotech
July 2018 – June 2019	**Board member of the Association of Doctoral Students in Life Sciences (ADSV)**
Institution	Swiss Federal Institute of Technology Lausanne (EPFL), Switzerland
Main activities	Managing communication and fostering scientific exchange between two EPFL campuses Co-organization of the 2nd PhD scientific retreat, May 2019, Ovronnaz, Switzerland
September 2013 – July 2014	**Member of eSTUDENT association**
Institution	University of Zagreb, Croatia
Main activities	Co-organization of Croatia's biggest student IT competition
July 2011 – July 2014	**Organizer of Lions Summer Youth Exchange Camps**
Institutions	Lions Clubs of Croatia
Main activity	Co-organization of a 2-week-long leisure camp for international students

Personal skills

Language skills	Croatian (mother tongue), English (C2), German (C2), Spanish (C1), French (B1)
Computer skills	
Programming Languages	C, CSS, C++, HTML, Java, JavaScript, MATLAB, PHP, Python, R, SQL, VHDL, XML, XSL
EEG analysis	FieldTrip, Cartool
Design of Application Specific Instruction Processors	DLX assembly language Tools: DLXsim, ASIP Meister, ModelSim, Xilinx ISE, CoSy compiler
Microcontroller programming	NXP LPC2103 with Keil MDK-ARM Version 5 IDE, TriCore
Printed Circuit Board (PCB) design	Altium Designer EDA
Gesture control devices	Myo, Kinect, Leap Motion
Product Lifecycle Management (PLM)	SolidWorks
Computer-aided Design (CAD)	SolidWorks
Computer Graphics	OpenGL, GLSL, Qt Framework
Virtual Reality platform	RealiSM
Other skills and competences	Karate master (black-belt) 1.st DAN, June 2010