

# Chapter 14

## Virtual Voice Illusions: A Short Review

**Ilusiones de voz virtual: una breve reseña**

Lisa E. Rombout

SCAN THIS QR CODE TO WATCH A VIDEO ABOUT THIS CHAPTER  
ESCANEE ESTE CODIGO QR PARA VER UN VÍDEO SOBRE ESTE CAPÍTULO

<https://www.youtube.com/watch?v=p-SVJvdN1eU&t=510s>



## Virtual Voice Illusions: A Short Review

Lisa E. Rombout

Department of Cognitive Science and Artificial Intelligence  
Tilburg University, The Netherlands  
[l.e.rombout@tilburguniversity.edu](mailto:l.e.rombout@tilburguniversity.edu)

### Abstract

Embodiment illusions show the flexibility of where our brain establishes the boundary between self and other. From rubber hands to full virtual bodies, multi-sensory integration (often combining vision with somatosensory or motor feedback) can lead us to experience artificial body-parts as if they are our own. Extended reality technologies allow us to experience embodiment over a wide range of human – and even some non-human – bodies, with subsequent effects on attitudes and behavior. Even our face, seemingly so closely linked to our identity, can undergo this effect through enfacement illusions. More recently, researchers have started developing and studying envoicement illusions as well. Sometimes described as the ‘auditory face’, the voice offers interesting novel territory for those interested in creating virtual embodied experiences. Additionally, illusions that are not dependent on vision can provide us with more information on the underlying mechanisms of embodiment. The findings from this new domain of ‘envoicement’ could influence how we approach embodiment in virtual applications, especially in virtual social situations.

**Keywords: embodiment, envoicement, virtual reality, voice illusions.**

## Introduction

Embodiment can perhaps be most succinctly described as the sense of having a body (Longo et al., 2008). This seemingly simple, subconscious process nonetheless consists of several complex subcomponents, such as agency (the sense that actions performed by a body are performed by you), ownership (the sense that things happening to a body are happening to you), and self-location (the sense of where the 'you' is located) (Kilteni, Groten, and Slater, 2012). Embodiment illusions have had a significant impact on our understanding of this ability to quickly distinguish between self and other from the myriad of incoming sensory signals.

The first embodiment illusion described in a scientific paper was the Rubber Hand Illusion (RHI) (Botvinick and Cohen, 1998). During the RHI, a participant's real hand is hidden from view with a screen, and a fake, rubber hand is placed where the real hand would be expected. Both real and fake hand are then synchronously stroked with a brush (at the same time and in the same place), while the participant watches. In under just a minute, a vast majority of these participants (Kalckert and Ehrsson, 2017) is overcome with a strong sense that the fake hand is in fact theirs. Of course, participants are consciously aware that this is not actually the case, but many instinctive reactions confirm the feeling; the proprioceptive sense of hand-location shifts away from the real hand towards the fake one (Rohde, Di Luca, and Ernst, 2011), and a sudden threat to the fake hand is physiologically processed as a threat to the person (Riemer et al., 2015).

The RHI has been repeated under many different conditions, often making use of virtual or augmented reality to expand the range of options for the appearance and behavior of the fake limb (IJsselsteijn, de Kort and Haans, 2006). High quality full body illusions have also become available with the advances of head-mounted displays, binocular cameras or completely virtual environments, and motion tracking technology (Petkova and Ehrsson, 2008). The combination of synchronous visual and motor feedback from a fully virtual body, or a remote-controlled robot, is able to establish a similarly strong feeling of embodiment. It is even possible to feel embodied with a virtual tail or a third hand (Steptoe, Steed, and Slater, 2013; Won et al., 2015).

The underlying process is likely to be some form of multi-sensory integration, where the brain receives all of the different sensory and feedback signals and through a process of causal inference, or predictive coding, arrives at a prediction of what is part of the body and what is not (Kilteni et al., 2015). The ready acceptance rate of most embodiment illusions shows that this process is rather flexible, both in terms of the time it takes to accept a different form, as well as the forms the brain will accept. Of course, the signals are not always in agreement – in the RHI, vision and somatosensation are telling a different story than proprioception. It is likely, however, that visual signals are weighed more heavily than others, with vision being the dominant human sense (Tsakiris, 2017).

Apart from the physiological effects, the features of the embodied body can also influence subsequent attitudes and behavior. For example, if light-skinned participants are embodied in a virtual avatar with a deeper skin tone, their implicit racial bias is significantly decreased afterwards (Banakou, Hanumanthu, and Slater, 2016). This effect occurs already after only a few minutes, even though implicit biases are normally quite difficult to change through cognitive effort. Similarly, the embodiment of adults in a child-like virtual body influenced their subsequent size-estimation of objects, and their attitudes to be more child-like (Banakou, Groten, and Slater, 2013). Enfacement, where a participant is embodied with a face in an augmented 'mirror', affects attributes such as self-recognition, mimicry and mood (Ma et al., 2016; Minio-Paluello et al., 2020; Tajadura-Jiménez, Grehl, and Tsakiris, 2012). This indicates that blurring between self and other can occur despite strong links between the body-part and self-identity, affecting associated attributes.

So far, we have discussed parts of the body that we can see and touch, or see and move, allowing these illusions to occur. However, there is another way we can act upon the world by moving our muscles, and that is by the production of sounds.

## Envoicement

The voice has been described as the 'auditory face', and is similarly linked to self-identity (Belin, Fecteau and Bedard, 2004). In recent years, several novel embodiment illusions acting on the voice have been created. This is not a straightforward task, as speaking is a very complex process that involves quick feedback loops on multiple levels (Postma, 2000). For any given utterance, feedback from multiple muscle groups in the thorax, throat and mouth areas is monitored to control the complex manipulation of the airstream that results in speech sounds. On top of that, auditory feedback is used to monitor the ongoing speech and make post-utterance corrections where necessary.

The first study explicitly exploring voice illusions in the context of embodiment was published in 2011 (Zheng et al., 2011). In this work, a stranger's voice was presented as the auditory feedback of a participant's own voice. As long as there was congruency between the utterance and the auditory feedback, the stranger's voice was experienced as the self, indicating that also on the vocal level bodily boundaries and identity can be flexible. Participants also altered their vocal characteristics in response to the illusion, although it was unclear whether this was a self-recognition, motor control, or mimicking response. Further study showed that, even though the voice illusion does not use vision as one of the contributing senses, a comparison with the RHI indicates that the perceptual inference that leads to either illusion is a stable trait within individuals (Zheng, Munhall and Johnsrude, 2016).

The idea of a voice illusion was subsequently picked up in a series of studies that explored whether the participant was actually required to speak themselves for the illusion to occur, or whether the effect could also be established with vibrotactile feedback on the throat. Most of these studies additionally combined the voice illusion with a virtual full body illusion, thus including the dominant visual feedback in what could be called a 'supporting' illusion. For example, it was shown that the feeling of agency over the virtual body, created chiefly by the integration of synchronous visual and somatosensory signals, could extend to a voice produced by that body (Banakou and Slater, 2014), although the effects on

subsequent changes in vocal characteristics varied (Banakou, and Slater, 2017). The earlier mentioned results regarding the embodiment of adults in a child-sized virtual body (Banakou, Groten and Slater, 2013) were also replicated with an added virtual voice (Tajadura-Jiménez et al., 2017). Although this addition did not alter the strength of the previously found effects, incongruency between the age of the virtual body and the virtual voice was found to disrupt the illusion, and intact illusions also influenced the vocal characteristics of the participant afterwards.

In a different study the virtual full body illusion was abandoned in favor of exploring auditory and vibrotactile feedback in isolation, showing that the lack of visual feedback makes the illusion much more difficult to establish (Rombout and Postma-Nilsenova, 2019). This is in line with the expectation that the dominance of visual feedback allows it to 'override' other senses more easily in the multisensory integration process (Tsakiris, 2017). A further comparison between articulated to non-articulated sounds, with the expectation that vibrotactile feedback on the throat matches the expected feedback of non-articulated sounds better, indicated that absence of feedback might actually disrupt the illusion less than the presence of asynchronous feedback (Rombout and Postma-Nilsenova, 2019). This is in line with predictive coding expectations (Kilteni et al., 2015) – the 'wrong' signal would cause an error, whereas the absent signal would not (Ainley et al., 2016).

It is clear that much remains to be explored in the realm of enoicement illusions. The complexity of feedback loops in the speaking process, and the relatively weak weighing of auditory signals as compared to vision, make the voice illusion a much more challenging candidate to establish than any primarily vision-based embodiment illusion. Supporting the enoicement with a simultaneous virtual body illusion could teach us more about the interplay between the two, whereas trying to establish a voice illusion without any visual support could potentially give us more insight in the predictive coding process that underlies these processes.

## Social Embodiment

Aside from explorations of the voice illusion itself, there is another realm where enoicement could make an interesting contribution. After all, the flexibility of body boundaries exemplified by embodiment illusions suggest a potential influence on our everyday interactions with other bodies. Synchronous behavior, such as dancing, marching or singing, is central to many human bonding experiences and affects interpersonal trust and closeness (Wiltermuth and Heath, 2009; Kreutz, 2014). Singing especially has been noted to have an ‘icebreaker effect’, increasing the feeling of emotional connection between people singing together (Pearce, Launay and Dunbar, 2015). Although most forms of joint speech in everyday life consists of dialogue (which is mostly not simultaneous), these effects of singing together suggest that there is perhaps a role for enoicement in social situations.

In one joint speech study, participants spoke simultaneously with another person (allowing reciprocal feedback) or with a recorded message, and either said the same thing or something else. When participants spoke synchronously at the same time as another person (meaning the motor feedback of the own voice would line up with the auditory feedback of the other voice), both utterances were processed by the brain as if they were other- instead of self-produced (Jasmin et al., 2016). This indicates a blurring of self-other boundaries reminiscent of the effect that the enoicement illusion has on identity – albeit not necessarily in the expected direction (Tajadura-Jiménez, Grehl and Tsakiris, 2012).

This result prompted more exploration of joint speech through the lens of embodiment. As suggested, joint speech can be both simultaneous (speaking at the same time) and synchronous (saying the same thing). A social enoicement study found that these influence speech adaptation differently. Synchronicity led to stronger pitch adaptation, potentially showing an embodiment effect similar to that found in voice illusion studies. Conversely, simultaneity led to suppression of adaptation in secondary vocal characteristics, perhaps indicating an ‘anti-embodiment’ effect, where subjects tried to maintain the body boundary between self

and other to avoid slurring of speech (Rombout and Postma, 2020). However, it is also possible that this negative shift actually represented a type of embodiment effect. Previous research has shown that alterations of the auditory feedback of the self-voice cause compensation behavior, shifting the voice away from the alteration (Jones and Munhall, 2000). Which explanation is correct remains to be seen.

## Conclusion

A lot remains to be discovered concerning the most functional design of enoicement illusions and their effects. These subtle non-visual illusions could potentially alter the lens through which we view joint speech interactions, or at least provide us with new insights. Additionally, they offer a novel sensory modality to explore in extended reality applications, especially those concerning virtual social situations and virtual bodies. Considering the virtual voice as a body-part opens up the possibility of influencing the behavior of the user through that avenue. The research in this area is very much still in its infancy, with plenty of avenues left to explore.

## References

- Ainley, V., Apps, M. A., Fotopoulou, A., & Tsakiris, M. (2016). 'Bodily precision': a predictive coding account of individual differences in interoceptive accuracy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160003.
- Alimardani, M., Nishio, S., & Ishiguro, H. (2013). Humanlike robot hands controlled by brain activity arouse illusion of ownership in operators. *Scientific Reports*, 3(1), 1-5.
- Banakou, D., & Slater, M. (2014). Body ownership causes illusory self-attribution of speaking and influences subsequent real speaking. *Proceedings of the National Academy of Sciences*, 111(49), 17678-17683.
- Banakou, D., & Slater, M. (2017). Embodiment in a virtual body that speaks produces agency over the speaking but does not necessarily influence subsequent real speaking. *Scientific Reports*, 7(1), 1-10.



- Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in Human Neuroscience, 10*, 601.
- Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences, 110*(31), 12846-12851.
- Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129-135.
- Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature, 391*(6669), 756-756.
- IJsselstein, W. A., de Kort, Y. A. W., & Haans, A. (2006). Is this my hand I see before me? The rubber hand illusion in reality, virtual reality, and mixed reality. *Presence: Teleoperators and Virtual Environments, 15*(4), 455-464.
- Jasmin, K. M., McGettigan, C., Agnew, Z. K., Lavan, N., Josephs, O., Cummins, F., & Scott, S. K. (2016). Cohesion and Joint Speech: Right Hemisphere Contributions to Synchronized Vocal Production. *Journal of Neuroscience, 36*(17), 4669-4680.
- Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F 0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America, 108*(3), 1246-1251.
- Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments, 21*(4), 373-387.
- Kalckert, A., & Ehrsson, H. H. (2017). The onset time of the ownership sensation in the moving rubber hand illusion. *Frontiers in Psychology, 8*, 344.
- Kilteni, K., Maselli, A., Kording, K. P., & Slater, M. (2015). Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in Human Neuroscience, 9*, 141.
- Kreutz, G. (2014). Does singing facilitate social bonding. *Music and Medicine, 6*(2), 51-60.
- Longo, M. R., Schüür, F., Kammers, M. P., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition, 107*(3), 978-998.
- Ma, K., Sellaro, R., Lippelt, D. P., & Hommel, B. (2016). Mood migration: How enfacing a smile makes you happier. *Cognition, 151*, 52-62.
- Minio-Paluello, I., Porciello, G., Gandolfo, M., Boukarras, S., & Aglioti, S. M. (2020). The enfacement illusion boosts facial mimicry. *Cortex, 123*, 113-123.
- Pearce, E., Launay, J., & Dunbar, R. I. (2015). The ice-breaker effect: Singing mediates fast social bonding. *Royal Society Open Science, 2*(10), 150221.

- Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: perceptual illusion of body swapping. *PLoS one*, 3(12), e3832.
- Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77(2), 97-132.
- Riemer, M., Bublatzky, F., Trojan, J., & Alpers, G. W. (2015). Defensive activation during the rubber hand illusion: Ownership versus proprioceptive drift. *Biological Psychology*, 109, 86-92.
- Rohde, M., Di Luca, M., & Ernst, M. O. (2011). The rubber hand illusion: feeling of ownership and proprioceptive drift do not go hand in hand. *PLoS one*, 6(6), e21659.
- Rombout, L. E., & Postma-Nilsenova, M. (2019). Exploring a voice illusion. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 711-717). IEEE.
- Rombout, L. E., & Postma, M. (2020). The One-Voice Expert. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society: Developing a Mind: Learning in Humans, Animals, and Machines* (pp. 3289-3294). Cognitive Science Society.
- Steptoe, W., Steed, A., & Slater, M. (2013). Human tails: ownership and control of extended humanoid avatars. *IEEE Transactions on Visualization and Computer Graphics*, 19(4), 583-590.
- Tajadura-Jiménez, A., Banakou, D., Bianchi-Berthouze, N., & Slater, M. (2017). Embodiment in a child-like talking virtual body influences object size perception, self-identification, and subsequent real speaking. *Scientific Reports*, 7(1), 1-12.
- Tajadura-Jiménez, A., Grehl, S., & Tsakiris, M. (2012). The other in me: interpersonal multisensory stimulation changes the mental representation of the self. *PLoS one*, 7(7), e40682.
- 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.
- Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science*, 20(1), 1-5.
- Won, A. S., Bailenson, J., Lee, J., & Lanier, J. (2015). Homuncular flexibility in virtual reality. *Journal of Computer-Mediated Communication*, 20(3), 241-259.
- Zheng, Z. Z., MacDonald, E. N., Munhall, K. G., & Johnsrude, I. S. (2011). Perceiving a stranger's voice as being one's own: A 'rubber voice' illusion?. *PLoS one*, 6(4), e18655.
- Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2016). A common perceptual inference for cross-modally induced illusions of body schema. *BioRxiv*, 066159.

## Ilusiones de voz virtual: una breve reseña

Lisa E. Rombout

Department of Cognitive Science and Artificial Intelligence  
Tilburg University, The Netherlands  
[l.e.rombout@tilburguniversity.edu](mailto:l.e.rombout@tilburguniversity.edu)

### Resumen

Las ilusiones de personificación muestran la flexibilidad de donde nuestro cerebro establece el límite entre uno mismo y el otro. Desde manos de goma hasta cuerpos virtuales completos, la integración multisensorial (que a menudo combina la visión con la retroalimentación sensorial o motora) puede llevarnos a experimentar las partes artificiales del cuerpo como si fueran nuestras. Las tecnologías de realidad extendida nos permiten experimentar la encarnación en una amplia gama de cuerpos humanos, e incluso algunos no humanos, con efectos posteriores en las actitudes y el comportamiento. Incluso nuestro rostro, aparentemente tan estrechamente vinculado a nuestra identidad, puede sufrir este efecto a través de ilusiones de encubrimiento. Más recientemente, los investigadores han comenzado a desarrollar y estudiar también las ilusiones de envío. A veces descrita como la “cara auditiva”, la voz ofrece un territorio novedoso interesante para aquellos interesados en crear experiencias corporales virtuales. Además, las ilusiones que no dependen de la visión pueden proporcionarnos más información sobre los mecanismos subyacentes de la encarnación. Los hallazgos de este nuevo dominio de “envoicement” podrían influir en cómo abordamos la encarnación en aplicaciones virtuales, especialmente en situaciones sociales virtuales.

**Palabras clave:** encarnación, envoicement, realidad virtual, ilusiones de voz.

## Introducción

Tal vez la encarnación pueda describirse de manera más sucinta como la sensación de tener un cuerpo (Longo et al., 2008). Sin embargo, este proceso subconsciente aparentemente simple consta de varios subcomponentes complejos, como la agencia (la sensación de que las acciones realizadas por un cuerpo las realiza usted), la propiedad (la sensación de que las cosas que le suceden a un cuerpo le están sucediendo a usted) y la autoconciencia. ubicación (el sentido de dónde se encuentra el “usted”) (Kilteni, Groten, and Slater, 2012). Las ilusiones de encarnación han tenido un impacto significativo en nuestra comprensión de esta capacidad para distinguir rápidamente entre uno mismo y el otro de la miríada de señales sensoriales entrantes.

La primera ilusión de encarnación descrita en un artículo científico fue la ilusión de la mano de goma (RHI) (Botvinick and Cohen, 1998). Durante el RHI, la mano real de un participante se oculta a la vista con una pantalla, y se coloca una mano de goma falsa donde se esperaría la mano real. Tanto la mano real como la falsa se acarician sincrónicamente con un pincel (al mismo tiempo y en el mismo lugar), mientras el participante observa. En menos de un minuto, la gran mayoría de estos participantes (Kalckert and Ehrsson, 2017) se sienten abrumados por una fuerte sensación de que la mano falsa es de hecho suya. Por supuesto, los participantes son conscientes de que este no es realmente el caso, pero muchas reacciones instintivas confirman el sentimiento; el sentido propioceptivo de la ubicación de la mano se desplaza de la mano real a la falsa (Rohde, Di Luca, and Ernst, 2011), y una amenaza repentina para la mano falsa se procesa fisiológicamente como una amenaza para la persona (Riemer et al., 2015).

El RHI se ha repetido en muchas condiciones diferentes, a menudo haciendo uso de la realidad virtual o aumentada para ampliar la gama de opciones para la apariencia y el comportamiento de la extremidad falsa (IJsselsteijn, de Kort and Haans, 2006). Las ilusiones de cuerpo completo de alta calidad también están disponibles con los avances de las pantallas montadas en la cabeza, las cámaras binoculares o entornos completamente virtuales y la tecnología de seguimiento de movimiento (Petkova and Ehrsson, 2008).

La combinación de retroalimentación visual y motora sincrónica de un cuerpo completamente virtual, o un robot controlado a distancia, es capaz de establecer una sensación de encarnación igualmente fuerte [9]. Incluso es posible sentirse encarnado con una cola virtual o una tercera mano (Steptoe, Steed, and Slater, 2013; Won et al., 2015).

Es probable que el proceso subyacente sea alguna forma de integración multisensorial, donde el cerebro recibe todas las diferentes señales sensoriales y de retroalimentación y, a través de un proceso de inferencia causal, o codificación predictiva, llega a una predicción de lo que es parte del cuerpo. y lo que no lo es (Kilteni et al., 2015). La rápida tasa de aceptación de la mayoría de las ilusiones corporales muestra que este proceso es bastante flexible, tanto en términos del tiempo que lleva aceptar una forma diferente como de las formas que aceptará el cerebro. Por supuesto, las señales no siempre están de acuerdo: en el RHI, la visión y la somato sensibilidad cuentan una historia diferente a la propiocepción. Sin embargo, es probable que las señales visuales tengan más peso que otras, siendo la visión el sentido humano dominante (Tsakiris, 2017).

Aparte de los efectos fisiológicos, las características del cuerpo encarnado también pueden influir en las actitudes y el comportamiento posteriores. Por ejemplo, si los participantes de piel clara están incorporados en un avatar virtual con un tono de piel más profundo, su sesgo racial implícito se reduce significativamente después (Banakou, Hanumanthu, and Slater, 2016). Este efecto ya se produce después de sólo unos minutos, aunque los sesgos implícitos suelen ser bastante difíciles de cambiar mediante el esfuerzo cognitivo. De manera similar, la encarnación de los adultos en un cuerpo virtual parecido a un niño influyó en su posterior estimación del tamaño de los objetos y en sus actitudes para que fueran más parecidas a las de un niño (Banakou, Groten, and Slater, 2013). Un participante se encarna con un rostro en un “espejo” aumentado, afecta atributos como el autorreconocimiento, la mímica y el estado de ánimo (Ma et al., 2016; Minio-Paluello et al., 2020; Tajadura-Jiménez, Grehl, and Tsakiris, 2012). Esto indica que la confusión entre uno mismo y los demás puede ocurrir a pesar de los fuertes vínculos entre la parte del cuerpo y la propia identidad, lo que afecta los atributos asociados.

Hasta ahora, hemos hablado de las partes del cuerpo que podemos ver y tocar, o ver y mover, permitiendo que ocurran estas ilusiones. Sin embargo, hay otra forma en que podemos actuar sobre el mundo moviendo nuestros músculos, y es mediante la producción de sonidos.

### Voz personificación

La voz ha sido descrita como la “cara auditiva” y está igualmente vinculada a la identidad propia (Belin, Fecteau and Bedard, 2004). En los últimos años, se han creado varias ilusiones de encarnación novedosas que actúan sobre la voz. Esta no es una tarea sencilla, ya que hablar es un proceso muy complejo que implica ciclos rápidos de retroalimentación en múltiples niveles (Postma, 2000). Para cualquier enunciado, se monitorea la retroalimentación de múltiples grupos de músculos en las áreas del tórax, la garganta y la boca para controlar la compleja manipulación de la corriente de aire que da como resultado los sonidos del habla. Además de eso, la retroalimentación auditiva se usa para monitorear el discurso en curso y hacer correcciones posteriores a la expresión cuando sea necesario.

El primer estudio que exploraba explícitamente las ilusiones de voz en el contexto de la encarnación se publicó en 2011 (Zheng et al., 2011). En este trabajo, la voz de un extraño se presentó como la retroalimentación auditiva de la propia voz de un participante. Siempre que haya congruencia entre el enunciado y la retroalimentación auditiva, la voz del extraño se experimentó como el yo, lo que indica que también en el nivel vocal, los límites corporales y la identidad pueden ser flexibles. Los participantes también alteraron sus características vocales en respuesta a la ilusión, aunque no estaba claro si se trataba de una respuesta de auto-reconocimiento, control motor o imitación. Un estudio adicional mostró que, aunque la ilusión de la voz no utiliza la visión como uno de los sentidos contribuyentes, una comparación con el RHI indica que la inferencia perceptiva que conduce a cualquiera de las ilusiones es un rasgo estable dentro de los individuos (Zheng, Munhall and Johnsruide, 2016).

La idea de una ilusión de voz se recogió posteriormente en una serie de estudios que exploraron si en realidad se requería que el participante hablará por sí mismo para que ocurriera la ilusión, o si el efecto también podría establecerse con retroalimentación vibrotáctil en la garganta. La mayoría de estos estudios combinaron adicionalmente la ilusión de la voz con una ilusión virtual de cuerpo completo, incluyendo así la retroalimentación visual dominante en lo que podría llamarse una ilusión de “apoyo”. Por ejemplo, se demostró que la sensación de agencia sobre el cuerpo virtual, creada principalmente por la integración de señales visuales y somatosensoriales sincrónicas, podría extenderse a una voz producida por ese cuerpo (Banakou and Slater, 2014), aunque los efectos sobre los cambios posteriores en las características vocales variado (Banakou, and Slater, 2017). Los resultados mencionados anteriormente con respecto a la incorporación de adultos en un cuerpo virtual del tamaño de un niño (Banakou, Groten and Slater, 2013) también se replicaron con una voz virtual adicional (Tajadura-Jiménez et al., 2017). Aunque esta adición no alteró la fuerza de los efectos encontrados anteriormente, se encontró que la incongruencia entre la edad del cuerpo virtual y la voz virtual interrumpió la ilusión, y las ilusiones intactas también influyeron en las características vocales del participante posteriormente.

En un estudio diferente, se abandonó la ilusión virtual de cuerpo completo en favor de explorar la retroalimentación auditiva y vibrotáctil de forma aislada, mostrando que la falta de retroalimentación visual hace que la ilusión sea mucho más difícil de establecer (Rombout and Postma-Nilsenova, 2019). Esto está en consonancia con la expectativa de que el predominio de la retroalimentación visual le permita “anular” otros sentidos más fácilmente en el proceso de integración multisensorial (Tsakiris, 2017). Una comparación adicional entre sonidos articulados y no articulados, con la expectativa de que la retroalimentación vibrotáctil en la garganta se corresponda mejor con la retroalimentación esperada de los sonidos no articulados, indicó que la ausencia de retroalimentación en realidad podría interrumpir la ilusión menos que la presencia de retroalimentación asincrónica (Rombout and Postma-Nilsenova, 2019). Esto está en consonancia con las expectativas de codificación predictiva (Kilteni et al., 2015): la señal “incorrecta” provocaría un error, mientras que la señal ausente no lo haría (Ainley et al., 2016).

Está claro que queda mucho por explorar en el ámbito de las ilusiones de envío. La complejidad de los bucles de retroalimentación en el proceso de habla y la ponderación relativamente débil de las señales auditivas en comparación con la visión, hacen que la ilusión de voz sea un candidato mucho más desafiante de establecer que cualquier ilusión de encarnación basada principalmente en la visión. Apoyar el envío con una ilusión de cuerpo virtual simultánea podría enseñarnos más sobre la interacción entre los dos, mientras que tratar de establecer una ilusión de voz sin ningún apoyo visual podría darnos más información sobre el proceso de codificación predictiva que subyace a estos procesos.

### Personificación social

Aparte de las exploraciones de la ilusión de la voz en sí, hay otro ámbito en el que el envío podría hacer una contribución interesante. Después de todo, la flexibilidad de los límites corporales ejemplificada por las ilusiones corporales sugiere una influencia potencial en nuestras interacciones cotidianas con otros cuerpos. El comportamiento sincrónico, como bailar, marchar o cantar, es fundamental para muchas experiencias de vinculación humana y afecta la confianza y la cercanía interpersonales (Wiltermuth and Heath, 2009; Kreutz, 2014). Se ha observado especialmente que cantar tiene un “efecto rompehielos”, aumentando la sensación de conexión emocional entre las personas que cantan juntas (Pearce, Launay and Dunbar, 2015). Aunque la mayoría de las formas de hablar en conjunto en la vida cotidiana consisten en diálogo (que en su mayoría no es simultáneo), estos efectos de cantar juntos sugieren que quizás exista un papel para la transmisión en situaciones sociales.

En un estudio de habla conjunto, los participantes hablaron simultáneamente con otra persona (permitiendo retroalimentación recíproca) o con un mensaje grabado, y dijeron lo mismo o algo más. Cuando los participantes hablaron sincrónicamente al mismo tiempo que otra persona (lo que significa que la retroalimentación motora de la propia voz se alinearía con la retroalimentación auditiva de la otra voz), ambas expresiones fueron procesadas por el cerebro como si fueran otras, en lugar de ellas mismas. producido (Jasmin et al., 2016). Esto indica una difuminación de los



límites entre el yo y el otro que recuerda el efecto que tiene la ilusión de enfrentamiento en la identidad, aunque no necesariamente en la dirección esperada (Tajadura-Jiménez, Grehl and Tsakiris, 2012)

Este resultado impulsó una mayor exploración del habla conjunta a través de la lente de la encarnación. Como se sugiere, el habla conjunta puede ser simultánea (hablando al mismo tiempo) y sincrónica (diciendo lo mismo). Un estudio de envío social encontró que estos influyen en la adaptación del habla de manera diferente. La sincronización condujo a una adaptación de tono más fuerte, mostrando potencialmente un efecto de encarnación similar al que se encuentra en los estudios de ilusión de voz. Por el contrario, la simultaneidad condujo a la supresión de la adaptación en las características vocales secundarias, tal vez indicando un efecto “anti-personificación”, donde los sujetos intentaron mantener el límite corporal entre ellos mismos y los demás para evitar la dificultad para hablar (Rombout and Postma, 2020). Sin embargo, también es posible que este cambio negativo representa realmente un tipo de efecto de realización. Investigaciones anteriores han demostrado que las alteraciones de la retroalimentación auditiva de la voz propia provocan un comportamiento de compensación, alejando la voz de la alteración (Jones and Munhall, 2000). Queda por ver cuál es la explicación correcta.

## Conclusión

Queda mucho por descubrir sobre el diseño más funcional de las ilusiones de envío y sus efectos. Estas sutiles ilusiones no visuales podrían alterar potencialmente la lente a través de la cual vemos las interacciones conjuntas del habla, o al menos proporcionarnos nuevos conocimientos. Además, ofrecen una novedosa modalidad sensorial para explorar en aplicaciones de realidad extendida, especialmente aquellas relacionadas con situaciones sociales virtuales y cuerpos virtuales. Considerar la voz virtual como una parte del cuerpo abre la posibilidad de influir en el comportamiento del usuario a través de esa vía. La investigación en esta área está todavía en su infancia, con muchas avenidas por explorar.

## Referencias

- Ainley, V., Apps, M. A., Fotopoulou, A., & Tsakiris, M. (2016). 'Bodily precision': a predictive coding account of individual differences in interoceptive accuracy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708), 20160003.
- Alimardani, M., Nishio, S., & Ishiguro, H. (2013). Humanlike robot hands controlled by brain activity arouse illusion of ownership in operators. *Scientific Reports*, 3(1), 1-5.
- Banakou, D., & Slater, M. (2014). Body ownership causes illusory self-attribution of speaking and influences subsequent real speaking. *Proceedings of the National Academy of Sciences*, 111(49), 17678-17683.
- Banakou, D., & Slater, M. (2017). Embodiment in a virtual body that speaks produces agency over the speaking but does not necessarily influence subsequent real speaking. *Scientific Reports*, 7(1), 1-10.
- Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in Human Neuroscience*, 10, 601.
- Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, 110(31), 12846-12851.
- Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3), 129-135.
- Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, 391(6669), 756-756.
- IJsselstein, W. A., de Kort, Y. A. W., & Haans, A. (2006). Is this my hand I see before me? The rubber hand illusion in reality, virtual reality, and mixed reality. *Presence: Teleoperators and Virtual Environments*, 15(4), 455-464.
- Jasmin, K. M., McGettigan, C., Agnew, Z. K., Lavan, N., Josephs, O., Cummins, F., & Scott, S. K. (2016). Cohesion and Joint Speech: Right Hemisphere Contributions to Synchronized Vocal Production. *Journal of Neuroscience*, 36(17), 4669-4680.
- Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F 0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, 108(3), 1246-1251.
- Kiltner, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4), 373-387.
- Kalckert, A., & Ehrsson, H. H. (2017). The onset time of the ownership sensation in the moving rubber hand illusion. *Frontiers in Psychology*, 8, 344.

- Kilteni, K., Maselli, A., Kording, K. P., & Slater, M. (2015). Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in Human Neuroscience*, 9, 141.
- Kreutz, G. (2014). Does singing facilitate social bonding. *Music and Medicine*, 6(2), 51-60.
- Longo, M. R., Schüür, F., Kammers, M. P., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition*, 107(3), 978-998.
- Ma, K., Sellaro, R., Lippelt, D. P., & Hommel, B. (2016). Mood migration: How enfacing a smile makes you happier. *Cognition*, 151, 52-62.
- Minio-Paluello, I., Porciello, G., Gandolfo, M., Boukarras, S., & Aglioti, S. M. (2020). The enfacement illusion boosts facial mimicry. *Cortex*, 123, 113-123.
- Pearce, E., Launay, J., & Dunbar, R. I. (2015). The ice-breaker effect: Singing mediates fast social bonding. *Royal Society Open Science*, 2(10), 150221.
- Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: perceptual illusion of body swapping. *PLoS one*, 3(12), e3832.
- Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77(2), 97-132.
- Riemer, M., Bublitzky, F., Trojan, J., & Alpers, G. W. (2015). Defensive activation during the rubber hand illusion: Ownership versus proprioceptive drift. *Biological Psychology*, 109, 86-92.
- Rohde, M., Di Luca, M., & Ernst, M. O. (2011). The rubber hand illusion: feeling of ownership and proprioceptive drift do not go hand in hand. *PLoS one*, 6(6), e21659.
- Rombout, L. E., & Postma-Nilsenova, M. (2019). Exploring a voice illusion. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 711-717). IEEE.
- Rombout, L. E., & Postma, M. (2020). The One-Voice Expert. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society: Developing a Mind: Learning in Humans, Animals, and Machines* (pp. 3289-3294). Cognitive Science Society.
- Stepto, W., Steed, A., & Slater, M. (2013). Human tails: ownership and control of extended humanoid avatars. *IEEE Transactions on Visualization and Computer Graphics*, 19(4), 583-590.
- Tajadura-Jiménez, A., Banakou, D., Bianchi-Berthouze, N., & Slater, M. (2017). Embodiment in a child-like talking virtual body influences object size perception, self-identification, and subsequent real speaking. *Scientific Reports*, 7(1), 1-12.

- Tajadura-Jiménez, A., Grehl, S., & Tsakiris, M. (2012). The other in me: interpersonal multisensory stimulation changes the mental representation of the self. *PLoS one*, 7(7), e40682.
- 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.
- Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science*, 20(1), 1-5.
- Won, A. S., Bailenson, J., Lee, J., & Lanier, J. (2015). Homuncular flexibility in virtual reality. *Journal of Computer-Mediated Communication*, 20(3), 241-259.
- Zheng, Z. Z., MacDonald, E. N., Munhall, K. G., & Johnsrude, I. S. (2011). Perceiving a stranger's voice as being one's own: A 'rubber voice' illusion?. *PLoS one*, 6(4), e18655.
- Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2016). A common perceptual inference for cross-modally induced illusions of body schema. *BioRxiv*, 066159.

## LISA E. ROMBOUT



### English

She is a human interested in quite a lot of things, including creative coding, cyborgs, virtual reality, regular reality, fairy-tales and the human brain. She is currently underlining this multidisciplinary madness by working towards a PhD in cognitive science and artificial intelligence, organizing creative coding meetups and hackathons, and teaching programming to children and adults.

### Español

Lisa E. Rombout es una humana interesada en muchas cosas, incluida la codificación creativa, los cyborgs, la realidad virtual, la realidad normal, los cuentos de hadas y el cerebro humano. Actualmente, está trabajando con locura en un proyecto multidisciplinar para obtener un doctorado en ciencia cognitiva e inteligencia artificial, organizando reuniones de codificación creativa y hackatones, y enseñando programación a niños y adultos.

