

Between Face and Voice: Semiotic Relationships

Original Study

Andrea Miccolis
Università degli Studi di Torino
andrea.miccolis@edu.unito.it

Received: 31 December 2024; Accepted: 1 August 2025

Abstract: The aim of this study is to investigate the semiotic relationships that emerge from the interpretation of the face in relation to the sound of the voice, considering both as signifying expressions subject to layered semantic interpretation. Specifically, we will explore the relationship between the voice and the face, treating them as palimpsests in symbiotic signification. This approach draws on fundamental semiotic research into the face, examining how, being both an integral part of the phonatory apparatus and a modulator of vocal sound, it conveys culturally legible characteristics and generates meanings in connection with the voice, particularly in relation to the emotional dimension perceptible in both.

Alongside this primary aim, a secondary objective emerges: to investigate, from an interdisciplinary perspective combining semiotics and cognitive science, the middle ground on which the contested boundary between nature and culture shifts. This reflection builds on analyses of the cultural interpretation of facial and vocal expression to reach the biological foundations of our semiotic system and its cultural ramifications.

Keywords: face, voice, palimpsest, mask, map

INTRODUCTION

Like a blind person moving through darkness, listening and drawing an inner representation of the territory they explore, this article attempts to describe the sign's relationships between voice and face. These elements are fundamentally intertwined within the continuity of the body and are therefore described as parts of a symbiotic system of sign palimpsests, encapsulating multiple layers of significations depending on the interpretative category applied.

A face can carry with it life stories, habits and emotions, a role also fulfilled by the voice. This work begins by exploring corporeality and the biological limits of these essential communicative systems. It then addresses simulation theory and physiognomy, assessing their strengths and weaknesses. Subsequently, we present

a brief interlude discussing an intriguing experiment by the Hungarian phonetician Iván Fónagy. Finally, we analyze a phenomenon from television entertainment, "The masked singer", drawing comparisons and contrasts with the theoretical framework established earlier and exploring the dimensions of identity and the mask, as they relate to this cultural product.

VOICE AS MAP OF CORPOREALITY

The voice and face are two fundamental elements of our corporeality involved in the communicative process. For this reason, when examining their involvement with meaning, it is essential to treat them as signifying apparatuses within a *semiotic-symbiotic system*. In such a system, the signification of the voice is enriched and complemented by that of the face and vice versa, forming

a continuous, mutual cross-reference that creates a complex and integrated system of meanings.

The voice, through variations in tone, timbre, rhythm and volume, is a powerful tool for communication. It conveys not only the verbal content of a message but also a wide range of paralinguistic information that influences how the message is interpreted. Through vocal modulation speakers can express emotions, intentions, moods and relational attitudes. For instance, intonation can transform a statement into a question or add an ironic or affectionate tone to an otherwise neutral sentence.

Similarly, the face is an equally potent vehicle of meaning. Facial expressions, eye movements, smiles and micro-expressions can communicate a wide range of emotions and moods in immediate and often unconscious ways. The face may reveal the sincerity of an emotion, express empathy, anger, surprise or disgust, often in ways that words cannot replicate. As noted earlier, in a semiotic-symbiotic system, voice and face do not function in isolation but continuously interact to generate and enrich meaning. For example, a smile can accompany a voice expressing joy, clarifying and authenticating the message. Conversely, a trembling voice might betray anxiety despite a calm facial expression, revealing underlying emotions. The mutual referencing between voice and face allows each to modulate the other's perception, creating a feedback loop that amplifies and clarifies the overall message. This interaction is especially evident in face-to-face communication, where all these elements contribute to constructing meaning and are filtered through the interpretative competencies of the interlocutors.

We must begin by considering the senses involved: sight, hearing and proprioception (also called kinesthesia), the latter being the sense through which we recognize the position of our body in space and the state of muscular contraction without the aid of vision.

According to Roland Barthes (1985, 239), "Listening is that preliminary attention that allows us to pick up anything that might alter the spatial system. It focuses on threat or need, and its material consists of the clues, be they signals of danger or promises of fulfillment"¹.

The concept of greatest relevance in this context is the mapping of the territory, understood as the environment in which signs are received. This means listening is not a passive but a dynamic process of interpreting signals from the surrounding world. Listening enables us to monitor and comprehend our environment by decoding signs, thereby preserving the integrity of the territorial system and responding to threats or opportunities.

During speaking, the sound produced is not merely the result of phonation; it also reflects subtle muscle

contractions in the face and mouth. These contractions shape the sound we produce and serve as tangible signs of how emotion is expressed through language, adding additional layers of meaning to words.

Given the profound connection among phonation, the body and emotions, it is helpful to divide the vocal phenomenon into three distinct phases: breath control, sound generation and sound modulation (Uberti 2005, 1). As phonation involves the entire body, three categories of interactions between body and voice can be identified: posture, the influence of the abdominal muscles on the larynx, and mimicry in emotional expression (*idem.*, 12).

Facial mimicry affects the color or quality of the voice. Contraction of facial muscles, during smiling, crying, or expressing astonishment, impact the laryngeal sound from its source. Facial movement also shapes the volume and contours of the vocal tract, resulting in signifying changes related to what is known as "resonance". Resonance is the process by which sound acquires texture and nuance through muscular modulation of bodily cavities. It may seem obvious, but it bears stating: our body does not produce sound in itself; it modulates a flow of air. It is the air that sounds, not us.

In vocal technique, physical attitudes such as smiling or yawning affect both biomechanics of the mouth and the vocal output, imparting emotional qualities even when induced by technique (Uberti 2005, 13). This reinforces the idea that the voice does not originate from the phonatory apparatus alone but is situated in the entire corporeality.

"Voice, corporeity of speech, is situated at the articulation of body and speech, and this is the place where the coming and going of speech will be able to take place"² (Barthes 1985, 247).

SIMULATION AND PHYSIOGNOMY

What has been discussed so far leads naturally into the framework of embodied cognition, which we explore further here to reach simulation theory. This theory posits an unconscious, mimetic process, often associated with the function of mirror neurons, through which we understand others by using our own mind as a model. We simulate beliefs, desires and other intentional states, projecting them onto others to interpret or predict their behaviors.

However, as Paolucci (2012, 252) rightly points out: "there is an obvious epistemological leap between saying that a neuron activates or does not activate and saying that a neuron simulates something"³. Nevertheless, there is an isotopy with the earlier discussion of body mapping and the mechanisms through which mirror neurons operate. According to Merlo (2013), who summarizes part of the theory of Rizzolatti and Sinigaglia (2006):

1 "L'ascolto è quell'attenzione preliminare che consente di captare tutto ciò che potrebbe alterare il sistema territoriale. Esso si focalizza sulla minaccia o sul bisogno, e il suo materiale è costituito dagli indizi, siano essi segnali di pericolo o promesse di appagamento". All translations from Italian are by Andrea Miccolis, unless otherwise noted.

2 "La voce, corporeità del parlare, si situa all'articolazione del corpo e del discorso, ed è questo il luogo dove potrà realizzarsi l'andirivieni del discorso"

3 "[...] c'è un evidente salto epistemologico tra il dire che un neurone si attiva o non si tratta e il dire che un neurone si attiva o non si attiva e il dire che un neurone simula qualcosa."



Figure 1. Facial expressions of the six basic emotions according to Ekman. Top row: anger, fear, disgust. Bottom row: surprise, happiness, sadness.

Mirror neurons can be activated in four ways: 1) by action, i.e., by the active first-person performance of a motor pattern, 2) by observation, 3) by hearing, e.g., hearing the sound of water activates the mirror system by making it perceive the image of a glass of water being filled, 4) by imaging a motor pattern without the need to actually perform it: e.g., if we think of a glass of water being filled⁴. (Merlo 2013, 3)

This perspective does not imply a deterministic interpretation or univocal interpretation of signs. Rather, we remain in the liminal space between nature and culture. The meanings attributed to signs are shaped by both biological constraints and cultural conditioning. As a result, some interpretations of these signs may vary significantly across cultures: what carries significance in one culture may lack symbolic value in another, and vice versa. This cultural variability underscores the complexity and relevance of the semiotic inquiry at hand.

A face or voice can reveal a lifestyle and emotional history, serving as palimpsest continuously rewritten by culture. Habits leave traces: aging alters appearance, voices become hoarse, drug use leaves marks; muscles may tighten or sag and body evolves. These changes tell stories of lived experience, inscribing choices onto the

body and lending significance to the indexes (or supposed indexes) read by culture.

Discussing such issues inevitably leads to physiognomy and its attempts to codify facial (and vocal) traits as indicators of specific passions or characteristics.

When linked with simulation theory, physiognomy not only offers a priori sign readings, such as Ekman's basic emotional facial expression⁵ (1992) (Fig.1), but also incorporates culturally shared emotional dialects, forming a neuro-cultural theory (Viola 2021, 56). This returns us to the central issue: subjectivity and identification. One may fall into the trap of internal recreation, responding emotionally to a sign because of personal and cultural overlays. Ekman's model, which leans more toward nature than culture, might be appropriate for a "raw" reading. Yet, cultural learning and individual experience, although embedded in culture, modify that framework. Thus, physiognomy may provide codes, but these cannot claim universality, as many once believed.

Earlier we addressed simulation, but we must now invoke its conceptual counterpart: dissimulation.

The distinction between "simulation" and "dissimulation" is a common denominator in the treatment of deception. This macro-distinction has served as the

⁴ "I neuroni specchio possono attivarsi in quattro modi: 1) grazie all'azione, cioè al compimento attivo in prima persona di un pattern motorio, 2) all'osservazione, 3) all'udito, ad esempio sentendo il rumore dell'acqua il sistema specchio si attiva facendo percepire l'immagine di un bicchiere che viene riempito, 4) all'immaginazione di un pattern motorio senza il bisogno di compierlo davvero: ad esempio, se pensiamo ad un bicchiere d'acqua che viene riempito".

⁵ While it is acknowledged that Ekman's theory has been partially outdated by more recent research, it remains heuristically valuable from a semiotic standpoint to consider it as a valid framework.

basis for discerning two major modes of deception, one oriented toward the logic of secrecy and the other toward revelation. The meaning of the Latin word *simulatio* [...] comes from *similis* ("similar") and means to make something like "fake" what it is not, to "pretend". Dissimulate, on the contrary, comes from *simulare* with the addition of the prefix and indicates an opposite and negative meaning. Dissimulate, in fact, means "to make something unrecognizable", "to hide", "to conceal" (Gramigna 2021, 130).⁶

Something may be concealed or made unrecognizable, yet concealment does not imply total erasure. Returning to Ekman: in attempting to suppress emotional facial expressions, one never fully succeeds. Traces and micro-expressions remain visible to the trained observer (Viola 2021, 56). If we accept the existence of micro-expressions on the face, then by extension we may consider the possibility of vocal micro-expressions as well.

FÓNAGY AND LA VIVE VOIX

A parenthesis must be opened here for the work of Iván Fónagy, a Hungarian phonetician who employed tools such as phonetic transcriptions, x-rays, spectrograms and other graphic aids in a fascinating study described in *La vive voix* (1983) (Valle 2019, 108). In this study, two groups of listeners evaluated two separate interpretations of Jean Cocteau's monologue *La Voix humaine* (1930), performed by two actresses.

The listeners were subsequently asked to fill out questionnaires regarding the perceived physical, moral and even social characteristics of the performers. The participants' responses showed non-random patterns; their answers aligned uniformly with each interpreter (Fónagy 1983, 161). The piece itself features a woman engaged in a phone conversation with a lover who has left her. His voice is never heard. The result is a genuine monologue.

Andrea Valle, in *Fisiognomica Udibile* (2019), referencing Fónagy (1983) adds a critical element that resonates with this study, the idea of "double encoding": "But [...] what happens to all the expressive aspects of voice that we also in all normalcy produce and receive? Hence the idea of a double encoding. On the side of the speaker, linguistic encoding is followed by a second, pathemic encoding: symmetrically, the receiver first decodes according to the pathemic model, then according to the linguistic model" (Valle 2019, 108).

Also worth noting is the "audible mimicry" experiment, in which an actress interprets a phrase with specific emotions. She is audio-visually recorded, and the

recordings are played for another actress who, without context, successfully imitates the original facial expression, underscoring the inseparable relationship between voice and face (Valle 2019, 109).

IDENTITY AND ABSENCE: THE MASKED SINGER

When a new territory is discovered and mapped, it is almost inevitably named, assigned an identity, which brings with it constraints rooted in cultural discourse.

The perception of a face is primarily based on its recognition. To recognize a face means to assign it to a class; and, in this respect, the perceptive experience seems already directed by the cultural filter: the filter of culture lies in between the face that is in front of us and our irresistible vocation to interpret⁷ (Magli 1995, 17).

A disfigured face or damaged vocal cords reduce the power to communicate with others and oneself. The absence or loss of these key identity traits emphasizes the essential role they play in defining selfhood, while also revealing that these bodily parts are worn by a primary subject who, in their absence, is stripped of all expression except bare, awkward existence: a garment we, in turn, wear. Despite the protective nuance of "garment", here we emphasize its aspect of covering and concealment, aligning it with the concept of the mask.

This section explores the mask as it relates to voice and identity in the reality singing competition show "The masked singer"⁸, the Italian adaptation of U.S. and German formats.

The show features celebrities competing in vocal performances while concealing their identities behind elaborate masks, usually depicting animals (Fig.2), either real or imaginary and full-body costumes. Their identities are revealed only after elimination. Until then, the participants remain isolated in private dressing rooms. The panel of judges and viewers at home attempt to identify them, based on vocal cues and performance.

The mask in this context recalls the tradition of theatrical masquerade, which Gramigna (2021) following Eco (1997 [1977]) explains clearly:

It is used not to deceive, but to pretend to be another. According to Eco, pretending is based on a convention; pretending is not a real form of deception but is a para-deception. Thus, pretending is not intended to deceive and therefore does not have the pragmatic dimension of deception. Pretending includes various

⁶ La distinzione tra "simulazione" e "dissimulazione" è un denominatore comune nella trattazione dell'inganno. Questa macro-distinzione è servita come base per discernere due grandi modalità di ingannare, una orientata alla logica della segretezza e l'altra alla rivelazione. Il significato del termine latino *simulatio* [...] deriva da *similis* ("simile") e significa rendere qualcosa di simile, "fingere" ciò che non è, "pretendere". Dissimulare, al contrario, deriva da *simulare* con l'aggiunta del prefisso e indica un significato contrario e negativo. Dissimulare, infatti, significa "rendere qualcosa irricognoscibile", "nascondere", "occultare".

⁷ "La percezione di un volto si basa innanzitutto sul suo riconoscimento. Riconoscere un volto significa assegnarlo a una classe; e, in questo senso, l'esperienza percettiva sembra già orientata dal filtro culturale: tra il volto che ci sta di fronte e la nostra irresistibile vocazione a interpretare [...] si interpone il filtro della cultura."

⁸ "Il cantante mascherato" in the Italian version. The concept is the same in other versions.



Figure 2. An illustrative example of the character masks employed in *The Masked Singer* television format.

phenomena, from theatrical masks to narrative fiction, counterfactuals, modeling simulations and protensive simulations (Gramigna 2021, 129).⁹

The crucial difference lies in the intention behind this mask. From others that may be considered functional to the artists wearing them and their creations. While some artists use masks to obscure or transform identity through makeup, prosthetics, digital filters or surgery, the masks in *The Masked Singer* are used purely for entertainment.

The absence of the gaze, which Nathalie Roelens (2021) identifies as a key factor in producing an uncanny or disturbing effect, is indeed a defining feature of the masks used in the show. However, in this specific context, that unsettling potential is neutralized. Rather than provoking discomfort or evoking a non-human dimension, the masks become part of a playful narrative that reorients the viewer's focus. Attention is not directed toward the mask as a surface or object, but rather toward the mystery it conceals, prompting an active engagement in uncovering the identity beneath.

This shift away from visual cues foregrounds the act of listening, not merely as auditory perception, but as an active interpretative practice. In the absence of access to the face, attention is redirected toward elements of voice and gesture. This form of listening seeks out subtle indicators: slips of the tongue, minor uncertainties in vocal timbre, rhythmic peculiarities and other distinguishing vocal features allowing a shift from ambiguity to relevance.

Beyond voice, certain body movements, especially those accompanying particular words or concepts,

also contribute to the decoding process. Such gestures are not random but embedded in individual expressive habits, which can offer further clues. The critical signs often emerge from slight divergences from regular patterns, revealing difference through variation. Yet, because the participants are public figures, the decoding depends on culturally specific knowledge, making this phenomenon non-universal from an anthropological standpoint.

What is even more interesting is the resemblance between this television format, centered on recognizing a concealed person, and Fónagy's (1983) experiment with Jean Cocteau's *La Voix humaine*.

The difference lies in practice: the former is academic, the latter, entertainment.

CONCLUSIONS

The themes developed in this work converge within a framework of simulation theory contextualized through semiotics embodied cognition. Through the synergistic relationship between vocal and facial signification, we understand them as a kind of cartographic metaphor: a map formed by interpreting clues, just as we interpret our environment and the objects within it.

By drawing on mirror neuron theory and embodied cognition, this study bridges two conceptual territories to revisit the debate over physiognomy from new angles. This framework may have its limitations, yet it resonates strongly, particularly when connected to the brief exploration of Fónagy's experiment in *La vive voix*.

In the final chapter, the case of *The Masked Singer* provided a compelling example to demonstrate how theoretical concepts can apply to real-world practices,

⁹ Essa non è usata per ingannare, ma per far finta di essere un altro. Secondo Eco la finzione fonda su una convenzione; il far finta non è una vera e propria forma di inganno, ma è un para-inganno. Quindi, il far finta non intende ingannare e pertanto non ha la dimensione pragmatica dell'inganno. Il far finta include fenomeni vari, dalle maschere teatrali alla finzione narrativa, dai controfattuali, alle simulazioni modellistiche e alle simulazioni pro-tensive.

while also extending the discussion of masks, identity and signification.

Of course, as a starting point, this analysis requires expansion to encompass the wide variety scenarios offered by semiotics and, especially, face studies. To conclude, let us recall a concept that has been fundamental in shaping this work from Marco Viola:

Working as if certain psychological theories were true, developing an analysis from those premises, and testing their validity. This entails the theoretical risk of assuming uncertain theories as true. However, this risk is the only alternative to silence; and since certainties are rare commodities in science, we might as well build on uncertainties by making the fragility of certain assumptions transparent.¹⁰ (Viola 2021, 64)

delle emozioni dalla fisiognomica alle emoji. Milano: Cinisello Balsamo, pp. 105–109.

Viola, M., 2021. Le espressioni facciali e i confini della semiotica. *Volti artificiali/artificial faces*. LEXIA. Rivista semiotica, 37–38, pp. 51–67.

REFERENCES

- Barthes, R., 1985. *L'ovvio e l'ottuso. Saggi critici III*. Torino: Einaudi.
- Cocteau, J., (2001 [1930]). *La voix humaine*. Roma: Stock.
- Eco, U., 1977. Dire il contrario, 35. In Bonfantini, M.A. (Ed.), 1997. *Menzogna e Simulazione*. Napoli: Edizioni Scientifiche Italiane, pp. 33–43
- Ekman, P. 1992. An Argument for Basic Emotions. *Cognition and Emotion*, 6, 3–4.
- Fónagy, I., 1983. *La vive voix: essais de psycho-phonétique*. Paris: Payot.
- Gramigna, Remo. 2021. Le forme della maschera: Aspetti semiotici della manipolazione del volto e della plasticità dell'apparenza. *Volti artificiali/artificial faces*. LEXIA. Rivista semiotica, 37–38, pp. 121–140.
- Merlo, G. 2013. *I neuroni specchio*. Università degli Studi di Genova. Available at: https://www.academia.edu/5835432/I_neuroni_specchio.
- Paolucci, C. 2012. Per una concezione strutturale della cognizione: semiotica e scienze cognitive tra embodiment ed estensione della mente. In Graziano, M., Luverà, C. (Eds), *Bioestetica, bioetica, biopolitica. I linguaggi delle scienze cognitive*. Messina: Corisco Edizioni, pp. 245–276.
- Magli, P., 1995. *Il volto e l'anima. Fisiognomica e passioni*. Milano: Edizione CDE.
- Rizzolatti, G., Sinigaglia, C., 2006. *So quel che fai. Il cervello che agisce e i neuroni specchio*. Milano: Cortina.
- Roelens, N., 2021. Animal Faces: The Question of the Gaze. *Volti artificiali/artificial faces*. LEXIA. Rivista semiotica, 37–38, pp. 29–50.
- Uberti, M., 2005. Acustica della voce. In Cingolani, S., Spagnolo, R. (Eds.), *Acustica musicale e architettonica*. Torino: UTET Libreria, ch.11.
- Valle, A., 2019. Fisiognomica udibile. In Pesenti Campagnoni, D., Arcagni, S. (Eds.), 2019. *Il volto*

¹⁰ "Lavorare come se certe teorie psicologiche fossero vere, sviluppando un'analisi a partire da quelle premesse, e verificarne la tenuta. Questo comporta il rischio teorico di assumere come vere delle teorie incerte. Tuttavia, questo rischio è l'unica alternativa al silenzio; e poiché le certezze sono merce rara nella scienza, tanto vale costruire sulle incertezze rendendo trasparente la fragilità di certe assunzioni".