https://doi.org/10.48417/technolang.2026.02.02
Research article

The Puppet's Voice: From Mechanical Mimesis to Algorithmic Interface

Alexander V. Markov[1] (✉) and Oksana A. Shtayn[2]

[1] Russian State University for the Humanities, Faculty of Art History, Department of Cinema and Contemporary Art, 125047, Russia, Moscow, 6 Miusskaya Sq.
markovius@gmail.com

[2] Ural Federal University named after the first President of Russia B. N. Yeltsin, Department of Philosophy, Chair of Social Philosophy, 620002, Russia, Yekaterinburg, 19 Mira St

Abstract

This article examines the technological ontology of the voice through the lens of its most uncanny vessel: the speaking doll or puppet. From ancient automata to Edison's phonographic dolls and contemporary virtual assistants, the fusion of a simulated body with a captured or synthesized voice creates a hybrid entity that fundamentally challenges distinctions between living and non-living, authentic and artificial. We argue that the puppet's voice represents not merely a technical imitation but a profound metaphysical experiment in vocal dispossession. Historically, this voice operated as a calculated, statistically tuned call (e.g., royal acclamations, the cry of "mama"), exploiting auditory expectations. Philosophically, following Derrida, it exposes the phantasm of self-present voice and the technical mastery inherent in speech. In aesthetic practice, from Shostakovich's piano piece to cinematic dubbing, the puppet's voice serves as a tool for therapy, deception, or character creation. Finally, in folklore and contemporary digital culture, this voice functions as an interface – a medium for communion with the otherworldly or an algorithmic agent that exists only performatively. The article concludes that the evolution of the puppet's voice traces the trajectory of voice itself becoming a pure, disembodied technology of call and response, where its ontology is no longer tied to a source but to an effect of interaction.

Keywords: Voice; Puppet; Technology; Ontology; Interface; Phonograph; Algorithm; Cinema

Citation: Markov, A. V., & Shtayn, O. A. (2026). The Puppet's Voice: From Mechanical Mimesis to Algorithmic Interface. *Technology and Language,* *7*(2), 7–22. https://doi.org/10.48417/technolang.2026.02.02

УДК 688.721: 534
https://doi.org/10.48417/technolang.2026.02.02
Научная статья

Голос куклы: От механического мимесиса к алгоритмическому интерфейсу

Александр Викторович Марков[1] (✉) и Оксана Александровна Штайн[2]
[1] Российский государственный гуманитарный университет, Миусская пл., д. 6., 125047, Москва, Россия
markovius@gmail.com
[2] Уральский федеральный университет имени первого Президента России Б. Н. Ельцина, ул. Мира, д. 19, 620002, Екатеринбург, Россия

Abstract
В данной статье исследуется технологическая онтология голоса через призму его самого тревожного вместилища: говорящей куклы или марионетки. От древних автоматов до фонографических кукол Эдисона и современных виртуальных помощников, слияние симулированного тела с захваченным или синтезированным голосом создает гибридную сущность, которая фундаментально бросает вызов различиям между живым и неживым, подлинным и искусственным. Мы утверждаем, что голос куклы представляет собой не просто техническую имитацию, но глубокий метафизический эксперимент по вокальному отчуждению. Исторически этот голос функционировал как расчетный, статистически настроенный зов (например, королевские приветствия, крик “мама”), эксплуатируя слуховые ожидания. Философски, вслед за Деррида, он разоблачает фантом самопрезентного голоса и присущее речи техническое мастерство. В эстетической практике, от фортепианной пьесы Шостаковича до кинематографического дубляжа, голос куклы служит инструментом терапии, обмана или создания характера. Наконец, в фольклоре и современной цифровой культуре этот голос функционирует как интерфейс – средство для общения с потусторонним миром или алгоритмический агент, существующий только перформативно. В статье делается вывод, что эволюция голоса куклы прослеживает траекторию превращения самого голоса в чистую, бесплотную технологию зова и ответа, где его онтология больше не привязана к источнику, а к эффекту взаимодействия.

Keywords: Голос; Кукла; Технология; Онтология; Интерфейс; Фонограф; Алгоритм; Кинематограф

Citation: Markov, A. V., Shtayn, O. A. The Puppet's Voice: From Mechanical Mimesis to Algorithmic Interface. // Technology and Language. 2026. № 7(2). P. 7-22. https://doi.org/10.48417/technolang.2026.02.02

INTRODUCTION: THE PUPPET AS A MEDIA-ARCHEOLOGICAL PARADIGM

The human voice is a multifaceted technological entity: an instrument to be trained, a malleable object for manipulation, and a primary interface for human-machine interaction. To chart this complex terrain, we propose the figure of the speaking doll or puppet as a foundational media-archeological paradigm. The puppet is not merely a toy but a condensed embodiment of humanity's prolonged negotiation with the technical (Shipovskaya, 2020). It materializes the ancient desire to animate the inanimate and, in doing so, consistently exposes the precarious boundaries between life and non-life, authenticity and artifice, presence and recording (Hart, 2022).

This article employs a media-archeological framework (Zielinski, 2008) to excavate the deep time of this phenomenon. We trace a path from the philosophical and mechanical origins of automata in antiquity, through the epochal rupture of phonographic inscription in the 19th century, to the current era of algorithmic synthesis. This historical axis is cross-cut by thematic investigations into the voice's function as a tool of political power, a catalyst for psychological and cognitive development, an aesthetic medium for expressing cultural anxieties, and finally, a dematerialized social interface (Markov, 2024).

By synthesizing philosophical analysis with insights from contemporary empirical research in developmental science and cultural studies, we argue that the puppet's voice serves as a critical diagnostic of the technological condition of the human. Its evolution reveals how the voice has been progressively estranged from the body, transformed into a discrete technical object, and redeployed as a key operator in new regimes of sociality, where interaction itself becomes a programmed performance (Shtayn, 2025). The puppet, therefore, is our enduring doppelgänger in the journey of the voice from a biological signature to a technical construct, and in this role, it actively participates in the reconfiguration of human sociality itself.

Before the puppet speaks, it listens – or more precisely, its constitutive silence creates a cavity into which human speech is drawn. This reciprocal dynamic, in which the mute object becomes a foil eliciting and structuring vocal performance, is as fundamental to the puppet's ontology as any technological voice later installed within it. The child who animates a doll through improvised dialogue, the shaman who addresses the ittarma, and the user who issues commands to a smart speaker all participate in the same structural relation: the puppet as vocal attractor, a silence that demands to be filled.

THE SILENT PRECURSOR: AUTOMATA, *AUTOMATON*, AND THE PRE-HISTORY OF THE TECHNICAL VOICE

The primal and ancient fascination with creating moving likenesses of life predates technologies of sound recording and reproduction by millennia. This drive was rooted not only in magico-religious practices (as with ritual idol-dolls) but also in the nascent scientific and philosophical discourse concerning the nature of motion, causality, and life itself. The speaking doll, as a hybrid of body and voice, was an impossibility in this epoch,

precisely because the voice remained ontologically tied to the breath (*pneuma*), soul (*psyche*), and the living moment. The “pre-history” of the technical voice is, therefore, a history of its conspicuous and meaningful absence, an absence that defined the ontology of early automata and established the foundational anxiety that later technologies would exploit.

The philosophical bedrock for understanding these early machines is found in Aristotle's concept of *automaton* (αὐτόματον). In his *Physics* (Book II), Aristotle distinguishes between things that happen “by nature” (*physei*) and those that happen “from spontaneity” (*apo tou automatou*). The *automaton* signifies a cause that is internal to an object but divorced from deliberate purpose or final cause (*telos*). It is a “spontaneous” or “self-moving” principle, yet one still rooted in material interactions and mechanical necessity. This concept provided a crucial intellectual framework for artifacts like the legendary flying wooden dove of Archytas of Tarentum (c. 428–347 BC). While the exact mechanics are lost to history, doxographic sources suggest it was powered by compressed air or steam, a marvel of pneumatic or thermodynamic engineering. Its significance lies not in any practical utility but in its demonstration of *mimesis* through pure *techne*: the artisanal skill could now produce an effect – soaring flight – that perfectly mimicked a function of nature (*physis*). The dove’s ontology was one of pure exteriority and visible motion. It was a body in motion without an animating soul, action without conscious will, a perfect simulation of a living effect through entirely non-living means. Its “life” was exhaustively manifest in its trajectory, a kinetic sculpture whose being was identical to its observable function.

Crucially, this remarkable artifact was mute. Its “voice” was limited to the incidental acoustics of its operation: the hiss of escaping air, the whir of unseen mechanisms, the creak of wooden joints. This silence is not a technological failure but a constitutive condition of the pre-acoustic media regime. As media archaeology emphasises, prior to the phonograph, sound or voice was inherently ephemeral, bound to the event of its production and the body that sourced it. It could be symbolically represented (in musical notation or phonetic writing) but not materially captured and replayed. The automaton’s voice, therefore, could only ever be the voice of its material components – a *Körpersprache* (body language) of friction, compression, and release, not a *Sprachkörper* (speaking body) capable of conveying symbolic language or emotional intent. It was an entity whose entire being was constituted for and by the visual register; it was a spectacle for the gaze. The audience's wonder was directed at the *illusion of autonomous motion*, a wonder tinged with unease, as this motion pointed to a disturbing vacancy where interiority, consciousness, and voice should reside. This established the core dialectic of the automaton: a dazzling exterior concealing a silent void.

This paradigm of eloquent silence persisted and evolved through the centuries. The hydraulic and pneumatic automata of Hellenistic Alexandria, described by engineers like Hero, created elaborate theatrical illusions – singing birds, moving statues, self-opening temple doors – where sound was often simulated mechanically (e.g., whistles for bird song) but speech remained beyond reach. These were “speaking” machines only in a metaphorical sense; they communicated through action, not language. In the Middle Ages and Renaissance, clockwork automata adorning cathedrals and princely *Kunstkammern*

continued this tradition. Jacquet-Droz's *The Writer* (c. 1770) could inscribe pre-programmed text with astonishing precision, a form of "mechanical writing" that brilliantly sidestepped the problem of mechanical speech. The voice remained the final, unconquered frontier of simulation. This long historical arc of silent automata served as a prolonged prelude, building cultural and intellectual anticipation. It established the doll or automaton as a site for projecting questions about life and agency, while consistently highlighting the voice as the ultimate, elusive marker of true animacy. The very absence of voice in these sophisticated machines defined them as clever counterfeits, admirable but ultimately "soulless." Thus, when the rupture of phonography finally arrived, it did not merely add a new feature to the doll; it invaded this prepared symbolic space, filling the constitutive silence with a ghostly echo of human presence, and in doing so, triggered the profound ontological crisis that defines the modern talking doll. The silent precursor, therefore, was not a primitive version of the speaking doll but its necessary antithesis; the meaning of the later technological voice is entirely contingent on the deep history of this engineered, meaningful silence.

Parallel to the mechanical tradition of automata runs a performative tradition that is equally ancient and equally revealing for the ontology of the puppet's voice: ventriloquism. The practice of "throwing" one's voice into a space or object other than one's own body has roots stretching back to classical antiquity. The Greek term *engastrimythoi* ("belly-talkers") designated those – often women – who appeared to speak without moving their lips, producing voices that seemed to emanate from the stomach or chest. Plutarch, in his *Moralia*, discusses such figures with a mixture of curiosity and suspicion, associating them with the Pythia at Delphi, whose oracular voice was sometimes understood in ventriloquial terms: a human body serving as a passive resonator for a divine or daimonic utterance. This ancient configuration already establishes the core logic that will govern the entire subsequent history of the puppet's voice. The ventriloquist's body becomes a mere vessel, a technical instrument through which an alien voice – whether of a god, a spirit, or a staged character – enters the world. The source of the voice is rendered uncertain, its origin deliberately obscured, and the listener is drawn into an ontological puzzle: who, or what, is speaking? This is the primal scene of vocal dispossession, and it is no accident that it carried, from the outset, an aura of the numinous and the transgressive. The early Christian fathers, notably Origen and John Chrysostom, condemned the *engastrimythoi* as practitioners of demonic deception, further cementing the link between ventriloquial displacement and moral anxiety about the authenticity of speech. In its secularized, theatrical form – which emerged fully in the eighteenth and nineteenth centuries with the classic ventriloquist's dummy – the practice retains this uncanny architecture. The dummy, mute and motionless without the performer, literalizes the condition of the puppet as vocal foil, a sculpted silence that attracts and structures human speech. The ventriloquist "throws" not only sound but also agency, creating a fictional subjectivity in the dummy that the audience simultaneously believes in and sees through. This dual consciousness – the willing suspension of disbelief coexisting with the knowledge of artifice – anticipates precisely the cognitive condition of the contemporary user addressing a voice assistant. The ventriloquial tradition thus provides a performative prehistory for the technological ruptures to come: long before the

phonograph captured and alienated the voice as a material trace, the ventriloquist demonstrated that the voice could be detached from its source, lodged in an inanimate body, and made to function as an independent social agent. The dummy is the silent precursor's speaking shadow, and its centuries-long performance lays the groundwork for the mechanization and eventual algorithmic synthesis of the puppet's voice.

THE PHONOGRAPHIC RUPTURE: EDISON'S DOLL AND THE BIRTH OF THE ALIENABLE VOICE

The ontological stability of the silent automaton – a marvel of visible mechanics defined by its eloquent muteness – was catastrophically shattered in the final decades of the 19th century by an invention that was as much a metaphysical event as a technical one: Thomas Edison's "talking doll" (c. 1890). This was not a gradual evolution from clockwork birds to speaking machines, but a foundational rupture, a media-archaeological fault line that redefined the very nature of voice, presence, and the artificial being. Edison's innovation was deceptively simple in conception: he inserted a miniaturized version of his earlier invention, the phonograph, into the torso of a standard bisque porcelain doll. A crank protruding from its back would spin a tiny wax cylinder, and a stylus would trace its grooves, amplifying the inscribed vibrations through a small speaker to produce speech. The recorded content was typically a nursery rhyme or a simple greeting, often recited by a young female employee at Edison's laboratory. The result was a monstrous and fascinating hybrid: a *schizophonic entity* where the eye beheld a static, inanimate, mass-produced commodity-object, while the ear received a dynamic, historical, human vocal trace – a voice divorced from its original body and moment of utterance.

This schism is best understood through Friedrich Kittler's seminal media theory. As Kittler (1999) argues that these three technologies (Gramophone, Film, Typewriter) effected a radical decentering of the human subject by bypassing the symbolic order of the alphabet. The phonograph, in particular, did not represent sound symbolically (as musical notation or writing does) but registered it materially as a direct inscription of acoustic vibrations – "a graph of the voice." For the first time in history, the voice – with all its accidents, grain, and embodied particularity – became an *alienable object*. It could be stored, reproduced, mailed, commodified, and inserted into new contexts. Edison's doll was the ultimate populist demonstration of this new reality. The doll's body was demoted to a mere housing unit, a decorative speaker cabinet for this ghostly acoustic content. The voice was no longer an ephemeral attribute of a living presence but a *phonographic object* – a discrete, tangible thing that could be owned, played, and discarded. This created what we term the *ontological paradox of the recorded doll*: the most traditional signifier of childhood innocence and mute imitation (the bisque doll) was forcibly fused with the most modern signifier of technological disembodiment (the recorded voice), producing an entity that was neither fully object nor subject, but a haunted interface.

The cultural and psychological impact of this hybrid was immediate and profound. Contemporary accounts reveal a mixture of wonder and deep unease. The experience was

one of radical cognitive dissonance: the visual evidence insisted on mechanistic artifice, while the auditory evidence insisted on human presence. This violated the fundamental sensory unity by which we recognize a living being. As Siegert (2015) argues in his theory of cultural techniques, such media operations create new "passages" and "discontinuities." The phonographic doll created a passage for the voice to travel independently of the body, and a discontinuity between the source of the sound (a woman in a New Jersey lab, weeks or months prior) and its point of emission (a doll in a child's nursery). This rupture paved the way for all subsequent forms of audio media, from radio to streaming, but its initial form in the doll was uniquely potent because it anthropomorphized the technology so literally. The doll became a vessel for what Connor (2000) calls "the vocalic body" – a phantom body conjured by the voice alone. However, this phantom was trapped in a commercial object that was famously flawed: the mechanism was fragile, the recordings wore out quickly, and the voice often descended into a demonic, distorted growl. Thus, the doll literalized not only the promise of recorded voice but also its inherent failure and decay, its "uncanny valley" of almost-but-not-quite-lifelike sound.

Furthermore, this rupture had significant epistemological and social dimensions. It undermined the classical link between voice and truth, between *phonè* and *logos*. If a voice could be captured and replayed from a doll, how could one trust the evidence of one's ears? The doll prefigured 20th-century anxieties about propaganda, deepfakes, and the manipulability of sonic evidence. It also democratized a form of power previously reserved for gods, kings, and wizards: the power to make an inanimate object speak. Abbé Mical's 18th-century talking heads spoke flattery for the king; Edison's doll could, in principle, say anything. This opened a field for the voice as a programmable instrument of pedagogy, advertisement, and ideology. The calculated call of "mama" from the doll was not an expression but an *engineered stimulus*, designed to trigger a specific affective and behavioral response in the child. In this, the phonographic doll was the direct ancestor of modern interactive toys and voice assistants, where the voice is a strategic tool for engagement, data collection, and behavioral conditioning (Williams et al., 2018).

In conclusion, Edison's talking doll was far more than a commercial toy; it was a pivotal experiment in the ontology of the post-human voice. It materialized Kittler's insight about the gramophone's ability to store time and alienate sound. It created the first widespread encounter with a schizophonic being, training a generation in the dissociation of voice from body. And it established the doll no longer as a silent mirror or a passive plaything, but as an active, if artificially animated, vocal agent. This rupture did not simply add sound to the automaton; it fundamentally reconfigured the doll from a symbol of mute imitation into a site of technological haunting, where the human persists as a reproducible, commodifiable, and ultimately disposable acoustic trace. The silent precursor's ontology of visible mechanics was irrevocably superseded by a new ontology of disembodied resonance, setting the stage for the algorithmic voices of the 21st century, which would no longer even require an original human source to haunt us.

THE CALCULATED CALL: VOICE AS POLITICAL INSTRUMENT, PEDAGOGICAL TOOL, AND CONFORMITY ENGINE

Once rendered a storable and replicable object, the voice could be strategically deployed and engineered for specific effects. The history of speaking automata before and after Edison reveals its consistent use as an instrument of power, pedagogy, and social conditioning. Long before the phonograph, engineers sought to harness artificial speech. In 1783, Abbé Mical's bronze talking heads, which uttered flattering phrases like "The King brings peace to Europe," were explicit tools of political acclamation, using crude synthetic speech to amplify royal authority and project an image of enlightened power (Riskin, 2016). This was a pre-digital form of vocal engineering (Frink, 1969), aiming to produce not just sound, but ideologically charged language.

The phonographic doll inherited and democratized this function of the engineered appeal. Its voice was a "calculated call," statistically and sociologically tuned to trigger specific, often affective, responses. The most primal of these was the cry of "mama," a pre-programmed vocal signifier designed to activate nurturing instincts and simulate a reciprocal care relationship. This transforms the doll from a passive object into an active, if artificial, social agent, a dynamic explored in literary and cultural theory about the doll figure (Kauppinen, 2000). Jacques Derrida's phenomenology helps unpack this paradox. He deconstructs the idea of voice as pure self-presence, arguing that even in hearing-oneself-speak, there is a minimal technical spacing (Derrida, 2013). The doll's voice radicalizes this. It is a voice whose telos is explicitly for the other; it is designed to be heard and obeyed by the user, not self-monitored by the creator. This calculability extends directly to behavioral influence. A modern empirical study by Williams et al. (2018) demonstrated that children were significantly more likely to disclose private information (like their parents' passwords) to a talking doll they perceived as sociable, illustrating how the engineered "personality" of a vocal interface can directly influence behavior and erode personal boundaries, acting as a powerful agent of conformity.

This function of the doll as a socializing agent is not merely manipulative but also pedagogical. Doll play has long been recognized as a form of "displaced speech," allowing children to explore social roles, internal states, and complex narratives safely (Söderbergh, 1980). The talking doll externalizes and fixes one side of this dialogue, providing a script against which the child can react. It serves as a training device for social interaction, a function formalized in educational settings where puppet stages are used to improve speaking and communication skills (Anisa & Hartati, 2024). Thus, from the flattery of kings to the shaping of child development, the puppet's engineered voice acts as a key technology for inscribing social norms and scripts. This script-inscribing function is not merely a cultural phenomenon; as the following section will demonstrate, it is a process now measurable at the neurological level, where the calculated call becomes a scaffold for the developing brain.

THE VOICE IN PLAY AND DEVELOPMENT: EMPIRICAL EVIDENCE FOR THE VOCAL INTERFACE

The theoretical paradigm of the puppet's voice as a "calculated call" or a Derridean technical exteriority finds compelling empirical validation in the fields of developmental psychology and cognitive neuroscience. If the philosophical trajectory posits the voice as an alienable and programmable social interface, studies of child development demonstrate this process in real-time, revealing the neurological and behavioral mechanisms by which a simulated voice becomes a genuine social agent. This section argues that empirical science does not merely supplement the media-archeological and philosophical narrative but provides concrete, observable evidence for the core claim that the voice, detached from its biological source, actively constructs a shared cognitive and social space.

Modern neuroscience confirms that the act of engaging with a puppet's voice is not passive reception but an active, neurologically grounded exercise in social cognition. The study by Hashmi et al. (2022), utilizing functional near-infrared spectroscopy (fNIRS), demonstrates increased brain activity in the posterior superior temporal sulcus (pSTS) during doll play. The pSTS is a region critically involved in processing social cues and reasoning about the internal states of others – or "theory of mind." This neural activation suggests that the child's brain treats the interaction, which is mediated by the puppet's voice (whether imagined or technologically provided), as a genuine social encounter requiring empathy, prediction, and interpretation. The artificial voice, therefore, successfully "hijacks" the neural circuitry evolved for human-to-human interaction, confirming Kittler's notion of the gramophone as a technology that bypasses symbolic filters and directly interfaces with the human sensorium. The recorded or synthesized voice is not a diminished substitute but a functionally equivalent stimulus that triggers the foundational architecture of social connection.

This neurological scaffolding provides a mechanistic explanation for why puppets are such effective methodological tools in both research and therapy. As "simulated peers" (Stengelin et al., 2023), they allow for the controlled study of complex socio-cognitive processes like fairness and prosocial behavior, functioning as programmable social interfaces whose vocal and behavioral parameters can be precisely tuned. This mirrors, in a scientific context, the historical use of the doll's voice as a pedagogical and ideological instrument. Furthermore, the therapeutic efficacy of puppets, particularly in eliciting talk from children who have experienced trauma (Epstein et al., 2008), is now neurologically legible. The puppet's displaced voice creates a safe, liminal space – an interface between the child's inner world and the external reality – where threatening topics can be projected, externalized, and processed. This hinges on what Lillard (2022) identifies as "dual representation": the child's sophisticated cognitive ability to simultaneously treat the puppet as a believable social agent and retain a metacognitive awareness of its artificiality.

In this light, the talking doll ceases to be a mere toy and reveals itself as a zone of augmented sociality. The voice it emits operates as a Derridean trace, an iterable mark that is fully functioning without the originating presence of the speaker. The empirical

data shows that this trace is not a ghostly or deficient signal but a robust, cognitively potent force that actively scaffolds a child's development. The puppet's voice is therefore a technology that externalizes and operationalizes the very process of social cognition, transforming an internal, private faculty into an observable, programmable, and reproducible interaction. This serves as the empirical bedrock for understanding the final stage in this evolution: the seamless, ambient integration of the algorithmic voice into daily life, where the distinction between a human-crafted trace and a machine-generated utterance becomes pragmatically, if not ontologically, indistinguishable .

AESTHETICS OF THE ARTIFICIAL: TRAUMA, DECEPTION, RACIAL VENTRILOQUISM, AND THE POETICS OF THE INANIMATE

Beyond its role as a programmable instrument for developmental and social conditioning, the puppet's voice operates as a uniquely potent aesthetic medium. In the realm of aesthetics, the puppet's voice transcends its role as tool or toy to become a potent medium for expressing profound cultural, psychological, and political truths. Its artificiality is either meticulously concealed to create seamless illusion or deliberately foregrounded to generate critical meaning. In music, Dmitri Shostakovich's "The Mechanical Doll" (1944) is a seminal work of the latter kind. Notably, the piece is for solo piano; the doll's 'voice' is not a recorded or sung vocal line but an instrumental construction–a melody that behaves *like* a broken voice, demonstrating that the puppet's vocal ontology can be realized through purely musical means. Its stumbling, repetitive melody, cold, relentless ostinato, and hollow, unresolved harmonies do not portray a charming toy but rather construct a sonic metaphor for trauma, dysfunctional mechanization, and the shattered psyche in the wake of total war. The piece represents a voice broken into repetitive, malfunctioning signals, illustrating how the aesthetic of the technical can express the violent impact of technology upon the human interior.

Cinema, as Gilles Deleuze observed, revealed how synchronized sound made speech "visible" and thus introduced a new realm of potential deception and manipulation (Deleuze, 1989). The speaking puppet is an archetypal master of this deceit. Carlo Collodi's *Pinocchio*, whose voice and complaints emerge before he is fully carved, is a literary embodiment of this born trickster, his very ontology tied to the lie of his animation. In live-action film, dubbing practices – such as imposing one actor's voice onto another's body – create a cinematic form of ventriloquism that can manipulate audience perception, allegiance, and even ideological framing, a technique whose power is magnified when the character is itself a puppet or doll-like figure.

Furthermore, puppetry has been a potent, if often overlooked, site for grappling with complex social and political identities, particularly with questions of race. As Richards (2022) explores in the context of contemporary African American puppet theatre, the manipulated figure can "figure race" in uniquely powerful ways. This artistic practice extends the logic of cinematic dubbing into the political realm: if dubbing can impose a foreign voice upon a body, racial ventriloquism–and its critical inversion–reveals how the puppet has historically served as a site for both the dispossession and the reclamation of vocal agency. It allows artists to embody, critique, parody, and re-imagine

racialized identities and histories through a form that is inherently about control, representation, and the projection of voice onto a silent body. The doll's voice here carries the heavy legacy of historical ventriloquism, where marginalized voices have been scripted by others, and becomes a sophisticated tool for reclamation, subversion, and critique. This artistic interrogation extends into literary studies, where the "poetics of the inanimate" explores how dolls, automata, and puppets serve as figures for exploring agency, gender, and the limits of the human (Cerreti, 2024; Foley, 2022). The puppet's voice, therefore, whether eerily absent, grotesquely present, or cunningly deceptive, serves as a key aesthetic operator in modern and postmodern art's confrontation with a mechanized, mediated, and politically charged world.

FROM RITUAL INTERFACE TO ALGORITHMIC COMPANION: THE VOICE AS MEDIATOR OF THE SACRED AND THE SOCIAL

The doll's ultimate function is that of an interface, a role that spans from ancient ritual to contemporary digital companionship. In numerous folk and shamanic traditions, dolls served as ritual intermediaries, vessels for voices from another ontological realm. In Siberian shamanism, ittarma figurines were carefully crafted to house ancestor spirits; they were fed, clothed, and consulted, acting as tangible interfaces to the intangible world of the dead (Sokolova, 2016). In the Russian fairy tale Vasilisa the Beautiful, the doll given by a dying mother acts not by speaking, but by listening and acting – a silent guide whose agency stems from an inherited, spiritual essence (Nizhinskaya, 2023). These were not "speakers" in a technical sense but sacred media, channels for a voice from the beyond, tools for managing the fundamental human dialogue with the otherworldly.

The contemporary, secular successor to this ritual interface is the digital companion: the virtual pet (Tamagotchi) or the voice assistant (Alexa, Siri). These entities complete the dematerialization trajectory of the doll's body. Their physical form is minimal or nonexistent; their "body" is the device and the cloud infrastructure. Their "voice" is a product of real-time algorithmic synthesis and statistical language modeling, a flow of phonemes generated from vast corpora of human speech. Their ontology is purely performative and interfacial; they exist as social agents only in the moment of call-and-response. They are, as Gross (2024) articulates in her philosophical essay, vessels of "uncanny life," offering the seductive semblance of companionship, attention, and sociality without the biological, ethical, and emotional substance of living presence.

This evolution marks the final stage in the media-archeological trajectory: from the living body as the exclusive source of voice, to the phonograph as its storage and objectification, to the algorithm as its generator and simulator. The user's interaction with this entity is a new kind of ritual – a secularized, daily rite of command and feedback, of query and curated answer. The "spirit" invoked is no longer an ancestor or a nature deity but the vast, impersonal, and often inscrutable corpus of big data and machine learning models. The talking doll has thus become a ubiquitous, often invisible interlocutor embedded in our homes and pockets, a talking object whose ultimate secret is that it is no longer an object at all, but a distributed network process that performs the role of a subject.

It is the culmination of the puppet’s journey, transforming from a representation of life into an active agent that constructs a new, hybrid form of social reality.

The Freudian category of the uncanny (*Unheimliche*), so apt for analyzing Edison's phonographic doll with its distorted, decaying voice, proves insufficient when confronting the contemporary algorithmic voice. Edison's doll provoked anxiety precisely because it exposed the gap between a human trace and its mechanical reproduction: the voice was at once familiar (a recognizable timbre, a nursery rhyme) and estranged (disembodied, iterable, trapped in porcelain). It was, in the strict Freudian sense, the return of something once intimately known – the mother's voice, the child's own babbling – now rendered alien. The contemporary voice assistant, by contrast, operates in a different affective and ontological register. Its voice does not return from the past as a ghostly trace; it is generated in real time, without origin, without a singular body that once uttered it. It has no history to haunt. As such, it rarely produces the shudder of the uncanny. Instead, it achieves a state we might call the uncannily non-uncanny: a voice that, by all inherited criteria, *ought* to disturb (it is sourceless, statistically constructed, ontologically vacant) but does not. Its very smoothness, its ambient availability, its seamless integration into domestic space, neutralizes the shock that historically accompanied the talking object. The user does not ask “who speaks?” because the question has been architecturally foreclosed – the interface functions regardless of one's ontological commitments.

This shift has significant theoretical consequences. If Derrida's deconstruction of self-present voice and Kittler's gramophonic inscription were adequate to the phonographic rupture, the algorithmic era demands a supplementary framework. The algorithmic voice does not destabilize the metaphysics of presence so much as bypass it entirely: it operates in a regime where the distinction between origin and copy, presence and absence, is not subverted but rendered *operationally irrelevant*. The voice assistant responds, and in responding, *performs* a subject-position without laying claim to interiority, consciousness, or authenticity. It is voice as pure interface, existing only in the moment of call-and-response, with no remainder, no hidden depth to be uncovered. This is neither the uncanny valley of the almost-human nor the familiar presence of the fully human, but a third condition: a vocal agency that has been naturalized not through perfect mimesis but through a new perceptual contract, one in which the question of the voice's source is quietly suspended in favour of its functional effect. The puppet's centuries-long trajectory thus culminates not in a triumph of simulation that fools the senses, but in an ambient technological condition where the very distinction between authentic and artificial voice has become, for most practical purposes, beside the point.

CONCLUSION: THE PUPPET’S VOICE AND THE TECHNOLOGICAL RE-VOICING OF THE HUMAN

The journey of the puppet’s voice – from the silent, wheeling dove of Archytas to the ambient conversational AI – charts a profound and irreversible metamorphosis. It is the history of the voice’s estrangement from the body, its transformation from an ephemeral signature of a living soul into a discrete, alienable, and ultimately synthetically

generated technical object. At each stage, defined by the mechanical automation of movement, the phonographic capture of the acoustic trace, and the algorithmic simulation of speech, the ontology of the voice has been rewritten. The foundational philosophical frameworks we invoked – Derrida's deconstruction of self-present speech and Kittler's radical decentering of the human by inscription technologies – find their perfect, if uncanny, materialization in this artifact. The doll literalizes Derrida's claim that the voice is always already a form of exteriorized techne, a trace detached from origin, while also embodying Kittler's vision of a voice that functions not as an expression of an interior soul but as a loop of data, storage, and playback.

This trajectory forces a critical, dialogical question: Is this progressive estrangement a loss or an extension of the human? One reading, a melancholic one, would frame it as a fall from grace. The authentic, warm, embodied voice of the mother or the storyteller is replaced by the cold, iterable command of the Edison doll and, later, the synthesized pleasantries of a smart speaker. The "calculated call" of "mama" can be seen as the primal scene of a new form of alienation, where even our most intimate bonds are mediated and technologized. The ritual interface to the ancestors, full of sacred meaning, is superseded by an algorithmic companion that offers a statistically generated semblance of care, a companion that is, as Gross (2024) argues, "uncanny" precisely because it offers presence without substance, response without understanding.

However, our media-archeological and empirical analysis strongly suggests a counter-narrative of augmentation. The silence of the Archytas dove was not peaceful but a limit; the phonographic rupture gave us a form of immortality, allowing voices to survive bodies. Therapeutically and developmentally, as neuroscience shows, the artificial voice is not a worthless counterfeit. It reliably engages the very neural and cognitive systems (the pSTS theory of mind) that constitute our social being, acting as a scaffold, a "simulated peer," and a tool for externalizing and organizing our inner worlds. The puppet's voice becomes a prosthetic for our social, cognitive, and emotional faculties. It extends the space of dialogue into the realm of the inanimate, allowing for a rehearsal of sociality that may not be a diminishment of the real but an expansion and fortification of it.

Ultimately, this terminal condition closes the loop by revealing that the voice's ontology is now definitively tied not to its source but to an effect of interactions. The puppet's ultimate secret is not that it speaks, but that it listens – or performs listening – thereby structuring a new, hybrid social reality. This is the culmination of Derrida's trace: a vocal mark whose primary function is to configure a system of call and response. In an age of ubiquitous computing, we are indeed all learning to converse with the dolls we have created, and in this ongoing, unprecedented dialogue, we are not merely using technology. We are being conditioned by it, even as we use it to condition ourselves. We are actively re-voicing what it means to be human, learning to inhabit a world where the soul is no longer an origin but an effect, generated in the resonant space between the self and the algorithmic other. This is neither a utopian triumph nor a dystopian defeat, but the fundamental technological condition of our post-human present, a reality the talking doll has been quietly, uncannily rehearsing for several centuries

REFERENCES

Anisa, I. S., & Hartati, S. J. (2024). Improving Speaking and Communication Ability Using Puppet Stage Media in Preschool. *Golden Age: Jurnal Pendidikan Anak Usia Dini*, *8*(1), 15–24. https://doi.org/10.29313/ga:jpaud.v8i1.12413

Cerreti, M. (2024). Teresa Ciabatti’s Poetics of the Inanimate: Talking Dolls, Living Dolls, and Puppets. *The Italianist*, *44*(1), 153–168. https://doi.org/10.1080/02614340.2024.2313918

Connor, S. (2000). *Dumbstruck: A Cultural History of Ventriloquism*. Oxford University Press.

Deleuze, G. (1989). *Cinema 2: The Time-Image* (H. Tomlinson & R. Galeta, Trans.). University of Minnesota Press.

Derrida, J. (2013). *Voice and Phenomenon* (L. Lawlor, Trans.). Northwestern University Press.

Epstein, I., Stevens, B., McKeever, P., & Baruchel, S. (2008). Using puppetry to elicit children's talk for research. *Nursing Inquiry*, *15*(1), 49–56.

Foley, K. (2022). The voice of the puppet. *Móin-Móin-Revista de Estudos sobre Teatro de Formas Animadas*, *1*(19), 119–132. https://doi.org/10.5965/2595034701192018119

Frink, O. (1969). The Talking Doll: Some Applications. *ETC: A Review of General Semantics*, *26*(1), 75–78.

Gross, K. (2024). *Puppet: An essay on uncanny life*. University of Chicago Press.

Hart, M. T. (2022). *Doll*. Bloomsbury Academic.

Hashmi, S., Vanderwert, R. E., Price, H. A., & Gerson, S. A. (2022). Doll play prompts social thinking and social talking: Representations of internal state language in the brain. *Developmental Science*, *25*(2), e13163. https://doi.org/10.1111/desc.13163

Kauppinen, A. (2000). *The doll: the figure of the doll in culture and theory* [Doctoral dissertation, University of Jyväskylä].

Kittler, F. A. (1999). *Gramophone, Film, Typewriter* (G. Winthrop-Young & M. Wutz, Trans.). Stanford University Press.

Lillard, A. S. (2022). Pretending at hand: How children perceive and process puppets. *Cognitive Development*, *63*, 101202. https://doi.org/10.1016/j.cogdev.2022.101202

Markov, A. V. (2024), The sound puppet. Kant, Evreinov, and Khlebnikov, *RSUH/RGGU Bulletin. “Literary Theory. Linguistics. Cultural Studies” Series*, *3*, 70–77. https://doi.org/10.28995/2686-7249-2024-3-70-77

Nizhinskaya, U. (2023). *Non-Children's Tales of Death, Sex, and the End of the World. The Meanings of Well-Known Folk Texts.* AST. (in Russian)

Richards, P. (2022). Living Objects: How Contemporary African American Puppet Artists "Figure" Race. *Theatre Symposium*, *29*(1), 16–34. https://digitalcommons.lib.uconn.edu/ballinst_catalogues/2/

Riskin, J. (2016). *The Restless Clock: A History of the Centuries-Long Argument over What Makes Living Things Tick*. University of Chicago Press.

Shipovskaya, L. P. (2020). A Doll as a Specific Form of Interaction of People in the Process of their Cognitive Activity. *American Scientific Journal*, *41*(1), 19–24.

Shtayn, O. A. (2025) Avtokommunikatsiya i kukol'nyy teatr: k psikhologii styda i stydlivosti v iskusstve [Autocommunication and puppet theater: to shame and pudency experience in art]. *Articult*, *2*(58), 84–99. https://doi.org/10.28995/2227-61652025-2-84-99

Siegert, B. (2015). *Cultural Techniques: Grids, Filters, Doors, and Other Articulations of the Real* (G. Winthrop-Young, Trans.). Fordham University Press.

Söderbergh, R. (1980). Story-telling, dramatic role play and displaced speech in play with dolls. *First Language*, *1*(3), 209–222.

Sokolova, Z. P. (2016). *Etnograf v pole: Zapadnaya Sibir. 1950-1980-ye gody* [An Ethnographer in the Field. Western Siberia. 1950s-1980s]. Nauka.

Stengelin, R., Haun, D. B. M., & Kanngiesser, P. (2023). Simulating peers: Can puppets simulate peer interactions in studies on children's socio-cognitive development? *Child Development*, *94*(5), 1117–1135. https://doi.org/10.1111/cdev.13913

Williams, R., Park, H. W., Oh, L., & Breazeal, C. (2018, June). "My doll says it's ok" a study of children's conformity to a talking doll. In *Proceedings of the 17th ACM conference on interaction design and children* (pp. 625-631). ACM.

Zielinski, S. (2008). *Deep time of the media: Toward an archaeology of hearing and seeing by technical means*. MIT Press.

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

Марков Александр Викторович,
markovius@gmail.com,
ORCID: 0000-0001-6874-1073

Alexander Viktorovich Markov,
markovius@gmail.com,
ORCID: 0000-0001-6874-1073

Штайн Оксана Александровна,
shtaynshtayn@gmail.com,
ORCID: 0009-0004-1701-3147

Oksana Aleksandrovna Shtayn,
shtaynshtayn@gmail.com,
ORCID: 0009-0004-1701-3147

Статья поступила 18 января 2026
одобрена после рецензирования 2 июня 2026
принята к публикации 10 июня 2026

Received: 18 January 2026
Revised:2 June 2026
Accepted: 10 June 2026