Published as *_Article* in *On_Culture: The Open Journal for the Study of Culture*
(ISSN 2366-4142)

# Siren Songs and Echo's Response: Towards a Media Theory of the Voice in the Light of Speech Synthesis

## Christoph Borbach

christoph.borbach@uni-siegen.de

Christoph Borbach is a research fellow at the research training group Locating Media at the University of Siegen, where he conducts a research project entitled "Zeitkanäle|Kanalzeiten" ("Time Channels|Channel Times") on the media history of the operationalization of delay time from a media archaeological perspective. Borbach studied Musicology, Media, and History at the Humboldt University of Berlin. His bachelor's thesis dealt with radio theories between ideology and media epistemology. For his master's thesis, he studied the media-technical implementations of echoes. His research interests include media theory of the voice, media archaeology of the echo, operationalization of the sonic, time-critical detection technologies and their visualization strategies, and the occult of media/media of the occult.

## Keywords

speech synthesis, phonography, voice theory, embodied voices, speaking machines, technotraumatic affects

## Publication Date

Issue 2, November 30, 2016

## How to Cite

Christoph Borbach. "Siren Songs and Echo's Response: Towards a Media Theory of the Voice in the Light of Speech Synthesis." *On_Culture: The Open Journal for the Study of Culture* 2 (2016). <http://geb.uni-giessen.de/geb/volltexte/2016/12354/>.

Permalink URL: <http://geb.uni-giessen.de/geb/volltexte/2016/12354/>

URN: <urn:nbn:de:hebis:26-opus-123545>

*On_Culture: The Open Journal for the Study of Culture*
Issue 2 (2016): The Nonhuman
www.on-culture.org
http://geb.uni-giessen.de/geb/volltexte/2016/12354/

# Siren Songs and Echo's Response: Towards a Media Theory of the Voice in the Light of Speech Synthesis

## _Abstract

In contrast to phonographical recording, storage, and reproduction of the voice, most media theories, especially prominent media theories of the human voice, neglected the aspect of synthesizing human-like voices by non-human means. This paper takes this lacuna as a starting point for an inquiry into the media theory of (non)human voices under the premise that the epistemological difference between techn(olog)ical voice production and its mere re-production is illuminated by the mythological motifs of the Sirens and Echo, respectively. Interestingly, the interconnection between terror and tempting nonhuman voices, which is implemented in the cultural imaginary through the Sirens' song, can be identified in the media history of speech synthesis, which challenges the idea(l) of the human voice as an anthropological constant. The main concerns here are to re-read the critique of Derrida's *Of Grammatology* and other theories of the human voice in the light of speech synthesis and show how the oft-used term 'disembodied voice' is inadequate when it comes to describing phonographical, radiophonic, and telephonic hearing situations.

## 1_Mythological Recursions: Echo and the Sirens

> We already have a foreboding that phonocentrism merges with the historical determination of the meaning of being in general as *presence* [...].[1]

Mankind's fascination with the human voice is as old as mankind itself. In two of the most prominent collections of ancient myths — Ovid's *Metamorphoses* and Homer's *Odyssey* — the human voice is a crucial topos. The specific role and epistemic-situational embeddedness of the (non)human voice in both myths-collections reveal its two main media techniques, namely, phonographic voice *re*production and synthetic voice production.[2]

In the *Metamorphoses*, Ovid writes about the meeting of Narcissus and Echo, who is still embodied. Echo, once a noisy and talkative nymph, has been punished by being able to only repeat words that have been said before. Therefore, her encounter with Narcissus can be described as a conversation with a recursive structure, consisting of his words and her replaying of his words in the same temporal pattern but with altered semantic meanings. In the words of Ovid, Echo is:

> A nymph whose way of talking was peculiar
> In that she could not start a conversation
> Nor fail to answer other people talking.
> Up to this time Echo still had a body,
> She was not merely voice. She liked to chatter,

> But had no power of speech except the power
> To answer in the words she last had heard.
> Juno had done this: when she went out looking
> For Jove on top of some nymph among the mountains,
> Echo would stall the goddess off by talking
> Until the nymphs had fled. Sooner or later
> Juno discovered this and said to Echo:
> "The tongue that made a fool of me will shortly
> Have shorter use, the voice be brief hereafter."
> Those were not idle words; now Echo always
> Says the last things she hears, and nothing further.[3]

The Sirens, on the other hand, are not restricted to reproduction. In the *Metamorphoses*, Ovid describes the Sirens as women who, as a result of their punishment, were endowed with feathers, wings, and birds' feet. Their tempting and beautiful voices enchant and attract passing seafarers. In other words, the "sweet-voiced Sirens"[4] are not human, but have human voices. In the *Odyssey* by Homer, a horrified Circe warns Odysseus about the Sirens' song:

> Your next land-fall will be upon the Sirens: and these craze the wits of every mortal who gets so far. If a man come on them unwittingly and lend ear to their Siren-voices, he will never again behold wife and little ones rising to greet him with bright faces when he comes home from sea. The thrilling song of the Sirens will steal his life away, as they sit singing in their plashet between high banks of mouldering skeletons which flutter with the rags of skin rotting upon the bones.[5]

When sailing near the Sirens' island, Odysseus plugs his sailors' ears with wax (the first historical mention of earplugs) and commands them to tie him to the mast of his ship. In this position, unable to move and follow the Sirens' invitation to visit them but able to hear their song, while the sailors are able to move but unable to hear, Odysseus became the only man to hear the Sirens' tempting song from a safe distance.

In a media-theoretical re-reading, the two myths reveal an opposition between voice *re*production — phonographic techniques — on the one hand, and technical media of voice *pro*duction — speech synthesis — on the other. At the same time, this opposition juxtaposes human with nonhuman speech. Although many analyses have emphasized the question of gender in the Sirens' episode, the more fundamental point is to recognize that the Sirens are nonhuman beings. The song of the Sirens is a chanted articulation by the nonhuman: synthetic speech *avant la lettre*.

In its literal sense, 'phonography' means sound- or voice-writing, as it combines the Greek *phōnē* (sound, voice) and *graphē* (writing). It refers not to writing *about* sound (as this essay does) but to the actual writing of sound *itself*, which gives sound its material, indexical, graphical, and, therefore, objective equivalent as time-invariant, fixed materiality. Édouard-Léon Scott de Martinville, the inventor of the phonautograph, an autonomous sound writer and the first practical apparatus for actually writing sound,[6] used such terminology as early as 1857 when describing the uniqueness of his machine as "[l]e son, aussi bien que la lumière, fournit à distance une image durable; la voix humaine s'écrit elle-même (*dans la langue propre à l'acoustique*, bien entendu) sur une couche sensible (...)."[7] The knowledge of such a non-arbitrary form of writing sound as *sound graphs* (as opposed to using the arbitrary alphabetical notation) provides a transition towards the *material* dimension of sound and, as such, constitutes a necessary precondition for phonography. These sound graphs, as auditory traces, may support Jacques Derrida's assumption that "[a]ll graphemes are of a testamentary essence,"[8] but the main fact remains: phonographic voices are vocal, non-alphabetical recordings with all the individuality of particular human voices which are indexical traces of human bodies.

With the advent of the phonograph, the human voice has been subjected to the operational logic of technical media that store speeches and voices as traces of *the real*. Since then, terms such as the 'exact repeatability' and hence the 'suspension from cultural historical time' apply to the human voice. Fundamentally, phonographic voices are characterized by the fact that they can only echo what was already said. In the "Song of Mr. Phonograph," the apparatus presents itself as such an echo (although this personification is a paradox in itself): "My name is Mister Phonograph and I'm not so very old. My father he's called Edison and I'm worth my weight in gold. The folks they just yell into my mouth and now I'm saying what's true: For just speak to me I'll speak it back and you'll see I can talk like you."[9] The song, which was commissioned by the Edison Speaking Phonograph Company and published in 1878, illustrates that, since the invention of the phonograph, echoes from the past can be replayed not just metaphorically but through technological means. It is thus no surprise that the American Edward Hill Amet marketed his own phonograph under the name Echophon from 1896 on, while Thomas Edison, on his visit to France in 1889, was celebrated not only as the man who "tamed the lightning" for his research on the light bulb but also as the man who "organized the echoes."[10]

The crucial aspect of the phonographical media technology is that the human being who originally spoke into the phonograph can be spatially or temporally absent when his or her voice is reproduced. It is an auditory characteristic that can be described with Raymond Murray Schafer's term "schizophonia," the splitting of electroacoustic reproduction of a sound from its source.[11] Likewise, Jacques Derrida described the aim of phonography as conserving spoken language, "making it function without the presence of the speaking subject."[12] Although this phenomenological perspective conceals the phonographical effect of what Vivian Sobchack named "re-presencing the past,"[13] Derrida points to the central underlying principle of phonography when it reproduces speech of a "speaking subject," that is, a human.

The singing of the Sirens and media techniques of speech synthesis are not restricted to mere reproduction in the way that Echo and phonography were. Speech synthesis produces human-like speech that originates from a mechanical or technological apparatus or, nowadays, from techno-mathematical software that can produce speech that no person uttered before. One common principle of speech synthesis is the so-called 'concatenative method,' which concatenates small units of pre-recorded speech and stores them in a database. An alternative to this method is so-called 'articulatory synthesis,' a technique that is based on a model of the human vocal tract — be it the modern, digital, and, therefore, computational simulation on a computer or the literally analog(ical) imitation of the human speech organs, such as Wolfgang von Kempelen's speaking machine from 1791 or Joseph Faber's speaking machine Euphonia, which was introduced in Vienna in 1840. The main difference between the former and the latter lies in the involvement or non-involvement of a human speaker. Whereas concatenative methods employ human language that is separated into small short-time recordings, articulatory synthesis produces human speech in a totally artificial way. In other words, human-like speech is created through thoroughly nonhuman means. In the context of digital speech synthesis and its software, it is thus obvious that what is written on a code level — the writing of complex algorithms — is not derived from speech (Ferdinand de Saussure's thesis). Rather, speech is subordinate to writing, which allows for a re-reading of Jacques Derrida's *Of Grammatology* regarding his explanation that speech also follows the logic of writing, i.e., the logic of signs.

Contemporary observers of the two types of speaking machines — the phonographic, on the one hand, and the actual speaking machines, such as Joseph Faber's, on the other — emphasized their disparity in quality. As early as 1878, the French Count Théodore Achille Louis du Moncel noted in his book *Le téléphone, le microphone et le phonographe* that "there is a great difference between the production and the reproduction of a sound, and a machine like the phonograph, adapted for the reproduction of sound, may differ essentially from a machine which really speaks."[14] Du Moncel further remarked that the *re*production of sound and even articulate sound may be very simple (which the phonographic principle actually was) whereas the *pro*duction of articulated sounds required a number of special organs analogous to the human 'speaking apparatus.' Although du Moncel's historical situatedness allowed him access only to knowledge of a speaking machine that *emulated* human speech (to use a term from computer science), his assessment also applies to digital speech synthesis.

Given that the emergence of technical media like the phonograph and cinematograph and their capacity to store, repeat, and manipulate the effects of human presence began to change the very parameters of the term 'presence' in the late 19th century,[15] it is essential to acknowledge that those parameters had to undergo a second shift of emphasis, which was related to technologies and techniques of speech synthesis. Speech synthesis bridges the alleged dichotomy of absence and presence, as synthetic voices constitute an acoustical presence of non-existing human bodies and therefore represent a purely nonhuman presence. Such an account appears oxymoronic; yet, it captures the crucial conceptual principle of synthetic voices. It is not surprising that these questions of absence and presence, being and nonbeing, can also be found on the content side of the history of speech synthesis. For instance, the five-minute-long, 7-inch vinyl disc entitled "Computer Speech: Hee Saw Dhuh Kaet (He Saw The Cat)," which Bell Telephone Laboratories made available for educational use in 1963, contains original recordings of synthetic speech, along with an explanation of Bell Labs' newly developed computer speech synthesis technique based on punched cards. These recordings deal recursively with the core question of speech synthesis when the computer quotes the famous "to be or not to be"-passage from William Shakespeare's "Hamlet."[16] This fundamental question of being/nonbeing can also be identified in early debates and questions on synthetic speech. Was it the voice of a human that could be heard from a figure or apparatus (via hidden speaking tubes or even hidden speaking humans), or was it a synthetic voice? In other words, was it a human or a nonhuman voice that could be heard; was it Echo or a Siren?

While the human voice is usually the essential and radical condition (or, more precisely and literally, articulation) of human presence, it is the synthetic voice that is most challenging to the human (it)self. Through synthetic speech, the voice as immediate security of the human speaker loses its one-to-one correlation with the human body. Voices are not restricted to the realm of the human and humanity anymore but can be produced through technical media. If the voice and the ability to speak are the phenomenon and ability that have been used to distinguish the human from the nonhuman, as, for example, Aristotle did with his postulate that a voice is the sound produced by a creature possessing a soul (*De Anima*, 420b5-6), then speech synthesis explodes this existential constitution of the voice. This puts Western logocentrism into the vexing position that nonhuman objects can also have a voice. To understand this, a short overview of the Derridean conceptualization of signs (more precisely, the deconstruction of phonocentrism) is helpful.

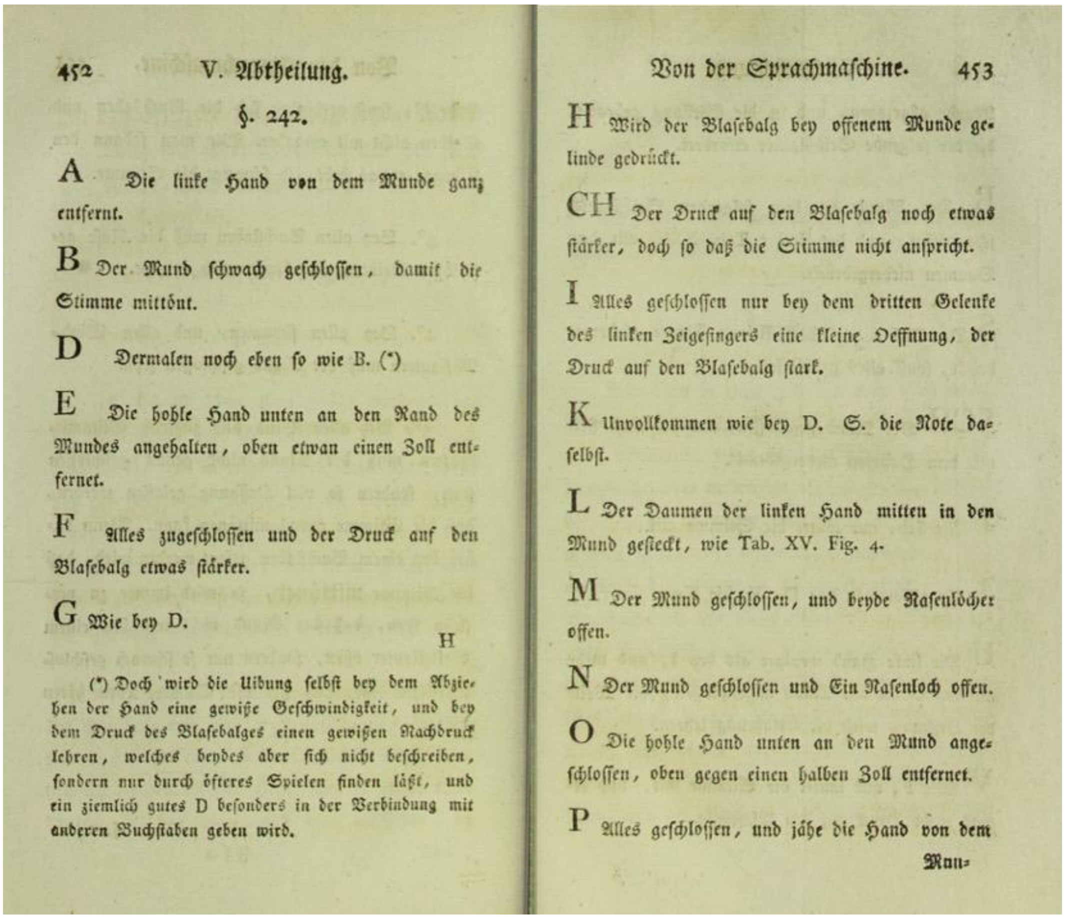
## 2_Phono-Logo-Centrism ≠ Voice as Message

For Derrida, the voice is more than just a medium for communication. Instead, it is the source of Western idea(l)s of truth, presence, and being. Derrida’s 1967 book on the neologism ‘grammatology’ — the knowledge of the letter, writing, and the written — is a critical re-reading of the entire Western philosophical tradition with the intention of considering the voice as a sign and aligning speech with the theory of signs, thereby overcoming oppositions dating back to antiquity. Derrida thereby relates Western phono- to logocentrism. ‘Logos,’ in a broad sense, means speech and its sense but also (the faculty of) reason. Therefore, logocentrism is “the belief that the first and last things are the Logos, the Word, the Divine Mind, the infinite understanding of God, an infinitely creative subjectivity, and, closer to our time, the selfpresence of full self-consciousness.”[17] According to Derrida, the originary presence of human beings is to be found in the human voice and the human ability to express mental experiences. Derrida traces this tradition of thought back to Aristotle, who referred to spoken words as symbols of mental experience and to written words as the symbols of spoken words. Put differently, written words are mere symbols of symbols and therefore supposed to be further away from human thought and actual presence.[18] Derrida also draws on Plato’s literary dialogue, the *Phaedrus*. In this dialogue, which can be identified as the first genuine media critique of writing, Plato argues that the written is a silent image of the spoken, just as paintings may have the appearance of liveliness but are in fact silent. Plato implicitly characterizes written signs as dead, whereas spoken, communicational interaction stays on the side of the living.

Responding to such theories about speech and writing, Derrida argues that vocal sound *also* functions and operates as a sign. This is because speech is potentially repeatable (its status *per se* is characterized by iterability, as Derrida points out), understandable, and 'readable,' even after the speaker's death. Vocal sounds are thus intrinsically quotable and able to be echoed, even before sounds could be technically reproduced. The Western notion and idea(l) of the immediacy of speech is thus a phantasm for Derrida. The main aim of his *Of Grammatology* is thus the "deconstruction of presence"[19] because "[i]mmediacy is [always] derived."[20]

In pursuing this line of argument, Derrida manages to avoid binary oppositions such as speech/writing, presence/absence, proximity/remoteness, outside/inside, meaning/representation, being/nonbeing (and, therefore, also human/nonhuman), and so forth. However, his argument deliberately conceals one dimension of the *phoné*, the vocal sound, that is to say that voice and speech are not equivalent, although they are deeply intertwined. Speech involves rhetorical, linguistic, and semantic categories whereas voice is the pure sound of speech, i.e., its materiality. The same difference holds between (technical) media and their content. Media are the precondition, the condition of possibility in the terms of Michel Foucault, the *arché* for communication (be it the Aristotelian *metaxy*, air, or complex digital computers). A genuine media-theoretical intervention can begin exactly at this point, following the maxim of Marshall McLuhan that it is not the content of a medium that is its message but the medium itself. The same fundamental difference between form and content applies to voice and speech. This reflection of the *material foundation* of communication and, therefore, the physics of speech is at the same time the advent of speech synthesis, as exemplified by the pioneering research of German physicist Christian Gottlieb Kratzenstein in 1779 on the production of vowel sounds.[21]

Although Derrida implicitly applies his theoretical analysis to speech's semantic categories, understanding his theoretical foundation of the relationship between speech and writing and their intertwining requires a re-reading in the light of digital speech synthesis. Digital speech synthesis is based on complex algorithms, which is to say writing, even though written code cannot be read as alphabetical language, as was the case at the time of Plato's *Phaedrus*. Plato argues that speech is prior to writing, but the opposite relationship applies in the case of digital speech synthesis: the hidden, written algorithmic structure precedes speech. The symbolism and symbolic character of speech is apparent when language is no longer indexically connected with a human being but is autonomized as a sign system and even based on the regulation of signs. This can be illustrated also by the analog sign instruction for the speaking machine of Wolfgang von Kempelen. In his book *Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine*[22] (*Mechanism of Human Speech Supplemented with a Description of a Speaking Machine*, no English translation), von Kempelen gives a set of instructions for his instrument-based auditory alphabet that requires manual operation (see Fig. 1). In this concrete situation, speech is not prior to signs. Rather, both rely on each other.



452 V. Abtheilung.

§. 242.

A Die linke Hand von dem Munde ganz entfernt.

B Der Mund ſchwach geſchloſſen, damit die Stimme mittönt.

D Dermalen noch eben ſo wie B. (*)

E Die hohle Hand unten an den Rand des Mundes angehalten, oben etwan einen Zoll entfernet.

F Alles zugeſchloſſen und der Druck auf den Blaſebalg etwas ſtärker.

G Wie bey D.

H

(*) Doch wird die Uibung ſelbſt bey dem Abziehen der Hand eine gewiße Geſchwindigkeit, und bey dem Druck des Blaſebalges einen gewißen Nachdruck lehren, welches beydes aber ſich nicht beſchreiben, ſondern nur durch öfteres Spielen finden läßt, und ein ziemlich gutes D beſonders in der Verbindung mit anderen Buchſtaben geben wird.

Von der Sprachmaſchine. 453

H Wird der Blaſebalg bey offenem Munde gelinde gedrückt.

CH Der Druck auf den Blaſebalg noch etwas ſtärker, doch ſo daß die Stimme nicht anſpricht.

I Alles geſchloſſen nur bey dem dritten Gelenke des linken Zeigefingers eine kleine Oeffnung, der Druck auf den Blaſebalg ſtark.

K Unvollkommen wie bey D. S. die Note daſelbſt.

L Der Daumen der linken Hand mitten in den Mund geſteckt, wie Tab. XV. Fig. 4.

M Der Mund geſchloſſen, und beyde Naſenlöcher offen.

N Der Mund geſchloſſen und Ein Naſenloch offen.

O Die hohle Hand unten an den Mund angeſchloſſen, oben gegen einen halben Zoll entfernet.

P Alles geſchloſſen, und jähe die Hand von dem

Mun-

Fig. 1: Instructions for perform speech with the speaking machine of von Kempelen: a short circuit between speaking and writing.[23]

Despite Derrida's interest in the voice beyond categories of communication, he nevertheless conceives the voice not as a materially and physically constituted entity but as a semantic system, in contrast to Mladen Dolar and his newer voice-theoretical approach.

As Dolar points out in his book *A Voice and Nothing More*, we usually fail to hear the voice and hear only the semantic meaning of the words spoken:

> If we speak in order to "make sense," to signify, to convey something, then the voice is the material support of bringing about meaning, yet it does not contribute to it itself. It is, rather, something like the vanishing mediator […] — it makes the utterance possible, but it disappears in it, it goes up in smoke in the meaning being produced.[24]

For Dolar, the voice is the material aspect of speech, the material carrier of meaning and sense, but it is not in itself semantic. Therefore, Dolar's minimalist definition of the voice in terms of its linguistic aspect is *that which does not contribute to making sense*,[25] a definition that follows a thesis that was proposed by Paul Zumthor in his essay "The Text and the Voice," in which he refers to the physical power of the human voice.[26]

*Prima facie*, the materiality of the voice seems to be merely a sub-category of its semantic and, therefore, communicational function as a medium of messages. But it is precisely the materiality of the voice, its existence not as a mere sign but as an *indication* in the sense of Edmund Husserl,[27] in contrast to the assumption of Derrida, that is the main feature of the individuality and uniqueness of human voices. This was also recognized by early radio theorists such as Rudolf Arnheim: "The pure sound in the word is the mother-earth from which the spoken work of art must never break loose, even when it disappears into the far heights of word-meaning."[28] This material quality, *le grain de la voix* (to use Roland Barthes' expression) is, therefore, a characteristic of the human voice that cannot be described by linguistics. Moreover, it is the irreducible individuality of a voice, its being something "beyond (or before) the meaning of the words, their form (the litany), the melisma, and even the style of execution: something which is directly the cantor's body, brought to your ears in one and the same movement from deep down in the cavities, the muscles, the membranes, the cartilages,"[29] or, in more technical terms, the hardware of the voice, which is at the same time — in Sybille Krämer's words — the "trace of the body when speaking."[30]

Barthes, in his essay "The Grain of the Voice," borrows Julia Kristeva's distinction between 'phenotext' and 'genotext' as a twofold opposition and modifies it to 'phenosong' and 'genosong' in order to describe the two realms of the voice: semantics and materiality.[31] Whereas 'phenosong' refers to the structure and code of speech, its semantic and invariant sphere, the 'genosong' refers to the "very materiality" of the voice, which has "nothing to do with communication,"[32] "something which is directly the [...] body."[33] This means that the 'grain of the voice' pertains to the interconnection between and intertwining of an individual voice and a human body. Therefore, the genosong serves as the theoretical model that permits the description of the affects and effects of synthetic speech, since it explicitly does not involve the content of speech but the unique sound of certain voices that endows them with individuality, as opposed to the mechanical sound of purely synthetic speech.

Like Barthes, Mladen Dolar is convinced that "[t]he voice without side-effects ceases to be a "normal" voice, it is deprived of the human touch that the voice adds to the arid machinery of the signifiers, threatening that humanity itself will merge with the mechanical iterability, and thus lose its footings." Therefore, it is "[p]aradoxically [...] the mechanical voice which confronts us with the object voice, its disturbing and uncanny nature, whereas the human touch helps us keep it at bay."[34] Focusing on such mechanical voices from the history of speech synthesis will show that Dolar's assumption is in line with contemporary witness reports on early speech synthesis. Upon hearing what Barthes calls "neutral voices" or "the whiteness of a voice," the listeners responded with a feeling of terror; in Barthes' words, "if sometimes that neutrality, that whiteness of the voice occurs, it terrifies us, as if we were to discover a frozen world, one in which desire was dead."[35]

## 3_Technotraumatic Voice Irritations[36]

In a posthumous note without any contextual information, Friedrich Nietzsche describes an imaginary moment of terror: "What I fear is not the horrible figure behind my chair but its voice: not even the words but the dreadful inarticulate and nonhuman sound of that figure. Yes, as if it were talking like people talk!"[37] Note that it is not the physical shape of the imaginary creature that frightens Nietzsche. The uncanny feeling and near-horror are triggered by its voice, the human-like voice of a nonhuman being, a voice from nowhere in the now-here. A report that is similar but was written about a real situation almost one hundred years earlier reads as follows:

> You cannot believe, my dear friend, how we were all seized by a magic feeling when we first heard the human voice and human speech which apparently didn't

> come from a human mouth. We looked at each other in silence and consternation and we all had goose-flesh produced by horror in the first moments.[38]

This note comes from a witness of the speaking machine of Wolfgang von Kempelen. The speaking machine consisted of a wooden box with a bellows (like a bagpipe) on one side, which served as lungs, and a rubber funnel that served as a mouth and had to be modified by hand to produce different vowels (see Fig. 2).

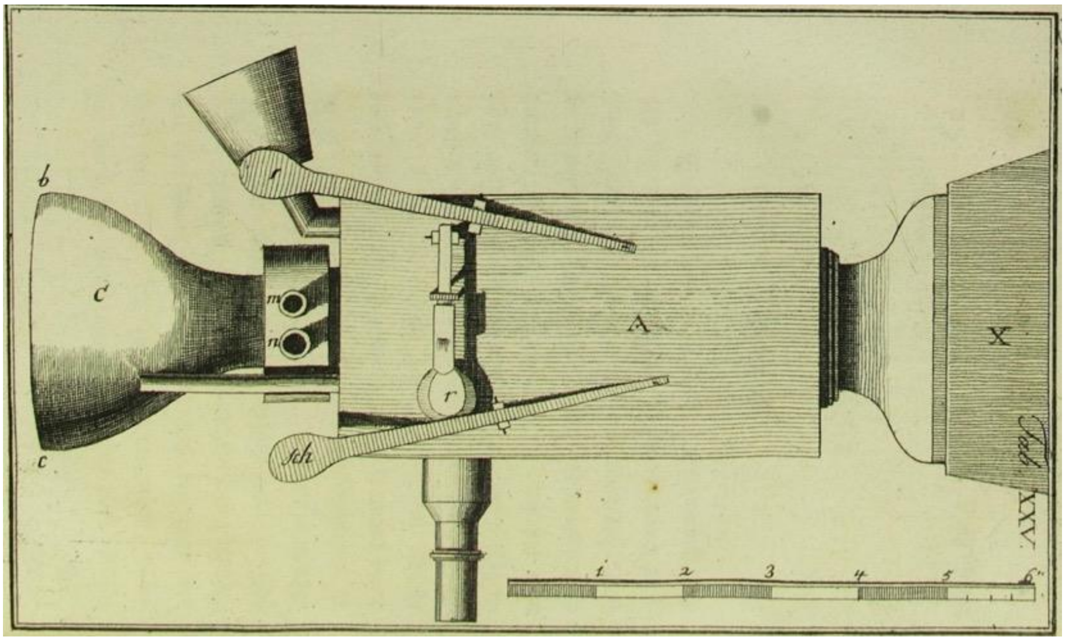


Fig. 2: The final speaking machine of Wolfgang von Kempelen; his own sketch.[39]

Another technologically similar mechanical speech synthesizer, named ‘Euphonia’ or the ‘Amazing Talking Machine,’ was presented by the German Joseph Faber in London in the Egyptian Hall around 1846. In the middle of the 19th century, this hall was commonly known as a place for occult demonstrations of automata that produce illusions. At the time, therefore, the exhibition hall was more associated with magic than with scientific instruments, which is what the Euphonia actually was in order to research human speech production. By pumping its bellows and manipulating a series of plates, chambers, and other apparatuses (including an artificial tongue, gums, and teeth), the operator could make Euphonia speak any European language (see Fig. 3).

The English journalist John Hollingshead attended one public demonstration of the machine and attested in his autobiography to the strangeness of the event, calling Faber’s Euphonia a “scientific *Frankenstein* monster”:

> In the centre [of the Egyptian Hall] was a box on a table, looking like a rough piano without legs and having two key-boards. This was surmounted by a half-

> length weird figure, rather bigger than a full-grown man, with an automaton head and face looking more mysteriously vacant than such faces usually look. Its mouth was large, and opened like the jaws of Gorgibuster in the pantomime, disclosing artificial gums, teeth, and all the organs of speech. [...] The Professor was not too clean, and his hair and beard sadly wanted the attention of a barber. I have no doubt that he slept in the same room as his figure — his scientific *Frankenstein* monster — and I felt the secret influence of an idea that the two were destined to live and die together. [...] He explained its action: it was not necessary to prove the absence of deception. One keyboard, touched by the Professor, produced words, which slowly and deliberately in a hoarse sepulchral voice came from the mouth of the figure, as if from the depths of a tomb. It wanted little imagination to make the very few visitors believe that the figure contained an imprisoned human — or half human — being, bound to speak slowly when tormented by the unseen power outside. […] As a crowning display, the head sang a sepulchral version of "God Save the Queen," which suggested inevitably, God save the inventor.[40]

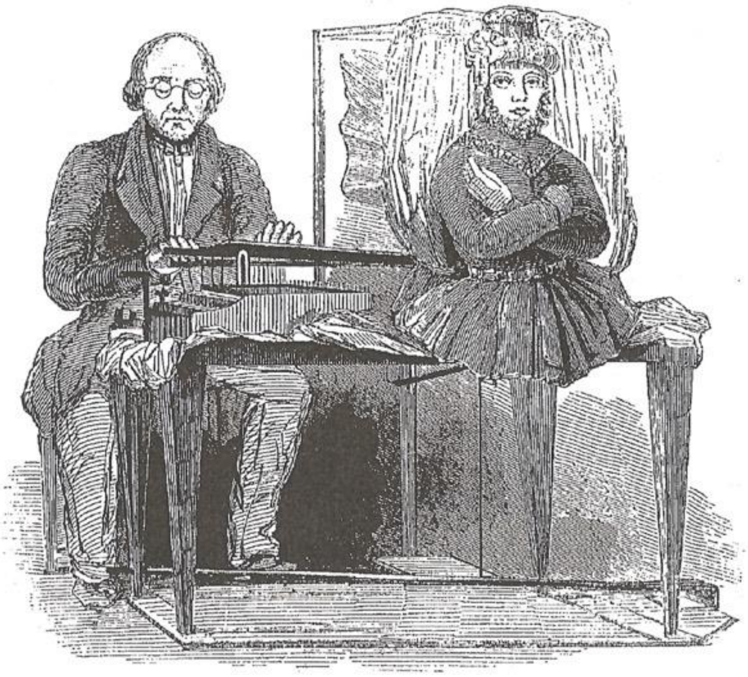


Fig. 3: Joseph Faber and his 'Euphonia' or 'Amazing Talking Machine.'[41]

Such rhetoric, with its references to death and the undead, the sepulchral, tombs, half-humans, and Frankenstein, was surely not coincidental but referred to the concrete articulation of human-like speech through nonhuman means. Euphonia, whose name came from the Greek prefix *eu-*, meaning well-, and *phōnē*, meaning voice, was anything but euphonious, or melodious, nor could the machine be described by the term 'euphonia' in the medical sense of the condition of having a normal, clear voice. Moreover, it was the first machine that realized the Sirens' song and recursively revealed its interconnecting of desire and death.

85 years after Faber's speaking machine, in 1931, a British journalist wrote in the London *Daily Express* after listening to a synthetic voice in London: "There was silence. The 'robot' voice had spoken. It was terrifying for the moment, almost horrible. I felt a tingle down my spine. I had heard a voice that was not a voice, words that had never been spoken."[42] The journalist was referring to experiments conducted by the British engineer and physicist Eric Allan Humphriss, who was working as a sound engineer for the British International Film Corporation at the time. The corporation had just completed a *sound* film starring the famous American actress Constance Bennett. Unfortunately, the movie's criminal character had the same name as a member of an aristocratic British family, who threatened to sue the film company. Since it was not possible to get the actress into the studio to re-record the dialogue with a different name for the character, a copy of the film was given to Humphriss, who manually re-worked Bennett's recorded voice in her physical absence. This was possible because the sound film was made with an optical recording process instead of synchronizing the images with a soundtrack recorded on a separate phonograph disc. This new optical recording technology used a microphone and a photosensitive selenium cell to translate sound waves into patterns of light. Those patterns were photochemically captured as small graphic traces on a strip that ran parallel to the celluloid film. Humphriss therefore had to analyze which wave patterns belonged to which sounds, which meant finding the correlating graphic images for all phonetic components. After doing so, he was able to arrange those components in a new order. For public demonstration, Humphriss arranged the word 'All-of-a-tremble,' now absolutely synthetic, on a forty-foot-long strip.[43]

Such techniques also allowed for the science fiction of re-vocalizing dead people in cases where recordings of their voices existed. For example, in a number of articles from the 1930s, the composer, music theorist, and journalist Arseny Avraamov proposed vocalizing the writings of Lenin, who had died in 1924, with the author's own voice, i.e. to synthesize his voice on the basis of his recorded speeches.[44] This can be read media-archaeologically as an attempt *avant la lettre* of text-to-speech synthesis. Speech synthesis therefore also includes the possibility of re-presencing the past as a sonic event that does not necessarily have to have an acoustic event as its original. That reveals another *horror* of synthetic voices, namely that they are the active speech of inanimate or dead things. While the advent of phonography allowed the dead to speak, such feats were limited to earlier recorded speech. The implications of synthetic voices go much further in that they resolve the paradox of Valdemar in Edgar Allan Poe's short story "The Facts in the Case of M. Valdemar," which arises when the protagonist remarks: "I *have been* sleeping — and now — now — *I am dead*."[45] What these experiential and fictional reports have in common and what they programmatically exemplify is that they are records of the subjective terror of nonhuman, neutral voices to which Roland Barthes referred.

Along with the term 'horror,' it is necessary to remember what Freud said in 1919 about the uncanny: "It is undoubtedly related to what is frightening — to what arouses dread and horror [...]."[46] In this essay, Freud refers to the German psychiatrist Ernst Jentsch and his 1906 article "On the Psychology of the Uncanny," in which Jentsch gives a short definition of the uncanny as something that is strangely familiar and as the uncertainty over whether an apparently animate being is really alive or, conversely, whether a lifeless object might in fact be animate.[47] To bolster his definition, Jentsch refers to literary works: "In storytelling, one of the most reliable artistic devices for producing uncanny effects easily is to leave the reader in uncertainty as to whether he has a human person or rather an automaton before him in the case of a particular character." And, "[t]his peculiar effect makes its appearance even more clear when imitations of the human form not only reach one's perception, but when on top of everything they appear to be united with certain bodily or mental functions."[48] This explanation has led to the modern application of Jentsch's psychoanalytical definition of the uncanny to the field of robotics and the perception of nonhuman beings or prostheses under the term "uncanny valley," which Masahiro Mori coined in 1970.[49] The 'uncanny valley' refers to the fact that things that look and move almost, but not exactly, like living beings cause revulsion among observers (see Fig. 4). I would like to argue that there is also an uncanny valley in the field of acoustics, specifically, in the field of speech and singing synthesis. The previously mentioned reports from the history of speech synthesis show this. Even nowadays, experiential reports implicitly address the uncanny valley.[50]

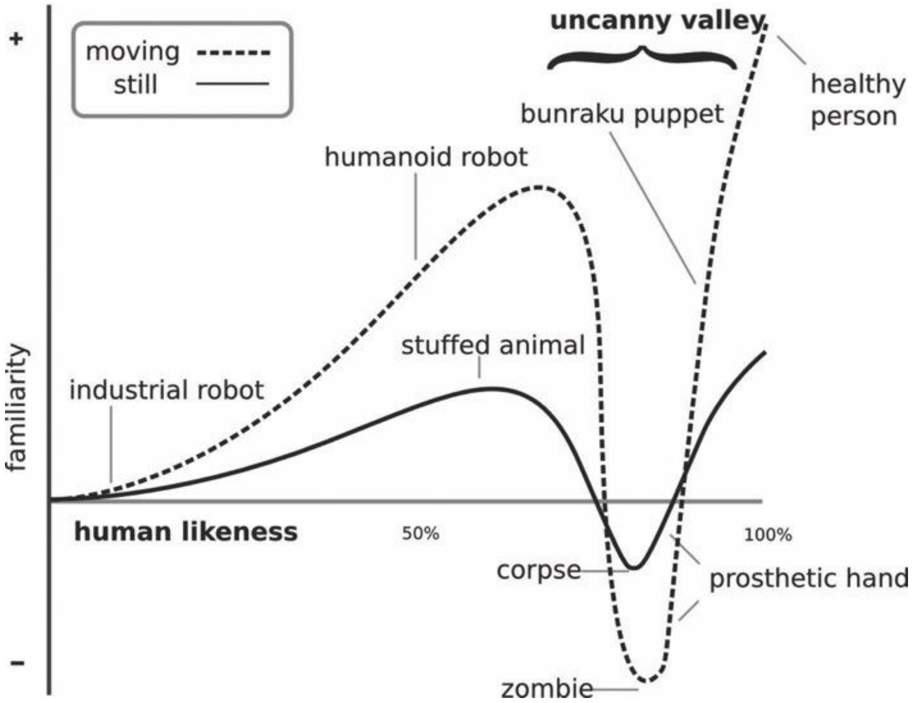


Fig. 4: The uncanny valley as described by Masahiro Mori in 1970.[51]

However, newer speech-synthesis software has developed to the point that it bridges the valley, causing positive responses among listeners. One programmatic example of this development is the Japanese singer Hatsune Miku, who is actually the *Vocaloid2* synthesizing software, which is a dynamic algorithm.[52]

## 4_Telephone Sex, or Corporealities — There are no Disembodied Voices[53] …

… except in speech synthesis and psychosis. Even the phonographic voice is always an embodied voice because it has a clear connection to, and an origin in one, and only one, physical body. Therefore, when we listen to voices, we do not merely hear voices; we also hear bodies. In this context, it is worth mentioning that the term ‘personare’ in the theater of ancient Greece means the listening to the actor’s voice through the theatrical mask,[54] which reveals the terminological and factual interweaving of the two terms ‘voice’ and ‘person.’

Although the term ‘disembodied voice’ is frequently used in media studies, communication studies, and related disciplines in the humanities, the fact that modern media have made the voice separable from the human body does not necessarily lead to the result that the separated voice itself is thereby *disembodied*. The ‘disembodied’ voice is a terminological trap, a theoretical assumption that does not stand up to the practices of hearing voice recordings and voice-only communication, as with the telephone, mobile telephone, voice chat software, and so forth. Some newer media theories borrow such phenomenological voice descriptions from the early years of radio theory without critically reflecting upon the terminology. As a consequence, most (media) theories of the human voice are self-contradictory.[55] I would therefore like to employ the term 'disembodied voice' as a merely *phenomenological* term, referring to the spatial or temporal absence of the speaker as a phenomenon of media-techn(olog)ically mediated voices. This proposal will be substantiated in the following with experiential reports and representative formulations from the early years of the phonograph and new neurological research and theories about the unity of voice and body.

The phonograph reminds its listeners of human presence, just as the dog Nipper, the trademark of the Gramophone Company, hears his master's actual voice, i.e., hears his presence. Such re-presencing of familiar voices had already been remarked upon in the earliest year of phonography. In an article in 1877, the *Scientific American* reported on the distinct reference of phonographic recordings to their producing hardware, that is, the genuine speaker with a known voice: "[C]ertainly nothing that can be conceived would be more likely to create the profoundest of sensations, to arouse the liveliest of human emotions, than once more to hear the *familiar* voices of the dead."[56] Hence, Derrida's argument, which I cited earlier, that the aim of phonography is to conserve spoken language, "making it function without the presence of the speaking subject,"[57] is superficial in that it deliberately conceals questions about reproducing presence through phonographic recordings.[58] In my criticism of Derrida, I do not intend to reintroduce binary oppositions such as speech/writing, but I aim to emphasize that recordings of the human voice are actually producing a kind of presence, even though the speaking subject is temporally or spatially absent — a presence of the absent. This is also the underlying main idea and explanation of the once-popular practice of phonoposting, the long-forgotten practice of mailing personal phonographical records, which was literally a practice and media-technique of voice mail *avant la lettre* in a media-archaeological sense.[59] Frederick Garbit's 1878 essay "Phonographic Letters," which explains that voices can be "identified" by the receiver of the envelope containing the phonographical cylinder, requires us to take the expression 'identified' literally, i.e., in the sense of a voiceprint;[60] a voiceprint, the spectrogram of a voice, allows the identification of a person via his or her voice biometrics. Just as every person has his or her own unique fingerprint, everyone has his or her own 'voiceprint,' a term — essential in the context of speaker recognition, identification, and authentication — that refers to the material sound aspects of a voice — just what the telephonic "it's me" indicates.

This identificational act goes hand in hand with an imaginary completion of the unity of voice and face. In the issue of *Scientific American* from February 12, 1887, the time of the advent of the telephone, an article appeared about the subjective feelings during the act of telephoning. The anonymous author explained that his telephonic shock was caused by a radically new experience, the confusion of "material non-existence" and "material existence"[61] triggered by the techno-revolutionary experience of hearing a person speak whose body is spatially absent. Nevertheless, the author explained, "I can *imagine* my friend at the other end of the line. But between us two there is an airy nowhere, inhabited by voices and nothing else — Helloland, I should call it."[62] In addition, Marshall McLuhan in his famous *Understanding Media* was convinced that "[a]s we read, we provide a sound track for the printed word; as we listen to the radio, we provide a visual accompaniment."[63]

The close connection and mutual interference of vocal sound and facial expression when articulating a certain sound has been known at least since the discovery of the McGurk effect in 1976. However, knowledge of the strength of the interconnection between the recognition of faces and voices is new. In 2011, researchers from the Max Planck Institute for Human Cognitive and Brain Science found a structural connection between voice and face recognition. Helen Blank and her team identified a direct interaction between voice- and face-processing areas of the human brain, leading to an interconnection between the recognition of voices and faces. Blank shows that, even if a person only hears another person talk, the area of the brain for face recognition is activated.[64] Blank's research could provide an explanation for the often irritating and challenging situation of combining a face with a voice that was previously only heard, as, for instance, when seeing a radio host whom one has before then only heard. This neural, imaginary completion of the voice-face-unity in the human brain also helps us to understand the affective power of voice recordings of familiar persons. It is a neurological manifestation of the media-archaeological thesis of re-presencing the past. Furthermore, this may explain why it is so irritating and even horrifying, to use the terms of Roland Barthes, to hear neutral, nonhuman synthetic voices: the main irritation comes not merely from the synthetic voice itself but from the neurological impossibility to imagine a human face for an apparently synthetic voice.

This confirms Theodor W. Adorno's assertion that "[t]he sound of any woman's voice on the telephone tells us whether the speaker is attractive."[65] Indeed, the practice of telephone sex is surely the best example to illustrate this paper's main thesis and programmatically outline its argument that the human voice, as the Lacanian *objet petit a*, is an object of (telephone-)sexual desire that refers to the physical materiality of the voice. Roland Barthes described this dimension with the term 'grain de la voix' and by borrowing the differentiation of phenotext and genotext from Julia Kristeva to describe the difference between semantic speech (phenotext) and non-semantic vocal sound (genotext), where the latter refers to the vocal identity and individuality of human voices that can be attractive. In the words of Roland Barthes, "[e]very relation to a voice is necessarily erotic, and this is why it is in the voice that music's difference is so apparent — its constraint to evaluate, to affirm" and "there is no human voice which is not an object of desire — or of repulsion,"[66] where the emphasis has to be put on the word 'human.' Consequently, the products of synthetic speech and singing fail to have individual, human *grain*, and are instead characterized through their uniform (and thus non-erotic) sound. This human-like speech that is actually nonhuman, as it is produced by instruments, media, or even software, is therefore a neutral voice in Barthes's sense, which produces a kind of affective horror, a claim that can be proven with reports of contemporary witnesses from the history of speech synthesis.

While Jacques Derrida's aim in *Of Grammatology* was to bring speech closer to the dimension of signs with his fundamental critique of logo- and phonocentrism, his argumentation can be re-read in the context of speech synthesis. However, it will not be until the advent of a computer-created synthetic voice that sounds naturally human, enabling listeners to imagine a face to go with the voice, that it will become impossible to distinguish whether one listens to a human or a computer voice. In conclusion, the process of deciding whether a voice is human or not (a kind of *acoustic* Turing test) is far more complex than we thought. It is not just a question of acoustic determination but also, as a matter of fact and not only of hypothesis, a question of neuropsychological facial imagination. Finally, and not only with telephone sex, the desire for a human voice is always also a desire for a human body, a body that does not exist in speech synthesis.

## _Endnotes

[1] Jacques Derrida, *Of Grammatology*, trans. Gayatri Chakravorty Spivak (Baltimore/London: The Johns Hopkins University Press, 1997 [1967]), 12.

[2] With the term ‘phonographic,’ I refer to all media techniques of recording, storing, and reproducing acoustical events and therefore also those of the human voice. This includes analog phonography and digital technologies as well.

[3] Ovid, *Metamorphoses*, trans. Rolfe Humphries (Bloomington: Indiana University Press, 1983 [1955]), 68. The Latin ‘reddere’ has been rendered as “answer” in this translation, but it can also mean to give back or return. Such a translation would conceive of words as signs or things that can be handed back or given away. This would strengthen the argumentation of the Derridean speech-as-sign-theory, to which I refer later on. I thank Cindy Heine for this observation.

[4] Ovid, *Metamorphoses* (cf. note 3), 116.

[5] T. E. Shaw, trans. *The Odyssey of Homer* (London: Humphrey Milford, 1935 [1932]), 170. Another deadly horror of the Sirens recurred during World War II. The Ju87 dive bomber of Nazi Germany’s Luftwaffe was equipped with sirens, so-called ‘Jericho trumpets,’ as an element of psychological warfare. Once the Ju87 started to dive, the sirens began to make a wailing noise, which later on became the typical sound motif of nosediving airplanes in movies.

[6] However, the *theoretical* idea of the material inscription of the human voice can be traced back at least to the mathematician and computer pioneer Charles Babbage. Cf. Charles Babbage, “The Ninth Bridgewater Treatise: A Fragment,” in *The Works of Charles Babbage* Vol. 9, ed. Martin Campbell-Kelly (London: William Pickering, 1989), 36.

[7] Patrick Feaster, ed. Édouard-Léon Scott de Martinville’s “Fixation Graphique de la Voice: A Critical Edition with English Translation and Facsimile,” Working Paper revised May 24, 2009, <http://www.firstsounds.org/publications/working-papers/First-Sounds-Working-Paper-03.pdf>; my emphases.

[8] Derrida, *Of Grammatology* (cf. note 1), 69.

[9] “The Song of Mister Phonograph,” Words & Music by H. A. H. von Ograph (New York: G. Schirmer, copyright 1878 by G. Schirmer). Today: Public Domain. The name of the author is definitely a pseudonym.

[10] Frank Lewis Dyer and Thomas Commerford Martin, *Edison: His Life and Inventions*. Vol. 2 (New York and London: Harper & Brothers, 1910), 746.

[11] Raymond Murray Schafer, *The New Soundscape: A Handbook for the Modern Music Teacher* (Toronto, Ont.: Bernadol Music Limited, 1969), 43–47.

[12] Derrida, *Of Grammatology* (cf. note 1), 10.

[13] Vivian Sobchack, “Media Archaeology and Re-presencing the Past,” in *Media Archeology: Approaches, Applications and Implications*, eds. Erkki Huhtamo and Jussa Parikka (Los Angeles: University of California Press, 2011), 323–333.

[14] Count Du Moncel, *The Telephone, the Microphone and the Phonograph* (New York: Harper & Brothers, 1879), 261.

[15] This is the central argument of the collaborative project *Archiving Presence: From Analog to Digital* of Humboldt University, Berlin, and Hebrew University, Jerusalem, between 09/2013–12/2015.

[16] “Computer Speech: Hee Saw Dhuh Kaet (He Saw the Cat)” [Vinyl]. Produced by Bell Telephone Laboratories, Incorporated, in 1963. Written and directed by D. H. VanLenten. This vinyl record

also features the famous synthetic singing of the IBM 704 of the song "Daisy Bell (Bicycle Built for Two)."

[17] Gayatri Chakravorty Spivak, translator's preface to Derrida, *Of Grammatology* (cf. note 1), ix–lxxxix, here: lxviii.

[18] Derrida recalls the Aristotelian definition as "Spoken words are the symbols of mental experience and written words are the symbols of spoken words." Derrida, *Of Grammatology* (cf. note 1), 30.

[19] Derrida, *Of Grammatology* (cf. note 1), 70.

[20] Derrida, *Of Grammatology* (cf. note 1), 157.

[21] For a general history of early synthetic speech research that also mentions Kratzenstein among other pioneers in the field, see Thomas L. Hankins and Robert J. Silverman, "Vox Mechanica: The History of Speaking Machines," in *Instruments and the Imagination*, eds. Thomas L. Hankins and Robert J. Silverman (Princeton: Princeton University Press, 2014), 178–220.

[22] Cf. Wolfgang von Kempelen, *Mechanismus der menschlichen Sprache nebst der Beschreibung seiner sprechenden Maschine* (Vienna: J. B. Degen, 1791).

[23] von Kempelen, *Mechanismus der menschlichen Sprache* (cf. note 22), 452–453.

[24] Mladen Dolar, *A Voice and Nothing More* (Cambridge, Mass.: The MIT Press, 2006), 15.

[25] Cf. Dolar, *A Voice and Nothing More* (cf. note 24), 15.

[26] The "*phōnē* does not take on meaning in any immediate fashion [...]." Cf. Paul Zumthor, "The Text and the Voice," in *New Literary History* 16.1 (1984), 67–92.

[27] Indications or indicative signs (*Anzeichen*), in contrast to expressions (*Ausdruck*), are not meaningful in themselves but refer to something *beyond* themselves, indicating a realm beyond their physical existence. Husserl called this the indicative relation of the sign and its significate. In the chapter "Expression and Meaning," Husserl defines the indicative signs as "certain objects or states of affairs *of whose reality someone has actual knowledge* indicate to him *the reality of certain other objects or states of affairs* [...]"; emphases in the original. The human voice is such an indicative sign, referring beyond its semantic meaning to the human presence, guaranteeing that a living being such as a speaker exists. See Edmund Husserl, "Expression and Meaning," in *The Essential Husserl: Basic Writings in Transcendental Phenomenology*, ed. Donn Welton (Bloomington and Indianapolis: Indiana University Press, 1999), 26–52, here 27.

[28] Rudolf Arnheim, *Radio*, trans. Margaret Ludwig and Herbert Read (New York: Da Capo, 1972 [1936]), 28.

[29] Arnheim, *Radio* (cf. note 28), 181.

[30] In the German original, Krämer states that "[d]ie Stimme ist die Spur des Körpers im Sprechen." Sybille Krämer, "Negative Semiologie der Stimme," in *Medien/Stimmen*, eds. Erika Linz and Cornelia Epping-Jäger (Cologne: DuMont, 2003), 65–82, here 67. Also in her numerous other papers about the voice, Krämer argues for its corporeal dimension; see, for example, Sybille Krämer, "Die 'Rehabilitierung der Stimme': Über die Oralität hinaus," in *Stimme: Annäherung an ein Phänomen*, eds. Doris Kolesch and Sybille Krämer (Frankfurt, Main: Suhrkamp, 2006), 269–295. In the subproject *Stimmen als Paradigmen des Performativen* (= *Voices as Paradigms of the Performative*), led by Krämer and Doris Kolesch, of the special research project *Kulturen des Performativen*, research into the role and performative embeddedness of the voice in (modern media) culture was funded between 2002 and 2010. For the approach of the project, see the programmatic paper Doris Kolesch, "Natürlich künstlich: Über die Stimme im Medienzeitalter," in *Kunst-Stimmen*, eds. Doris Kolesch and Jenny Schrödl (Berlin: Theater der Zeit, 2004), 19–38.

[31] Cf. Roland Barthes, “The Grain of the Voice,” in *Image Music Text: Essays selected and translated by Stephen Heath* (London: Fontana Press, 1977), 179–189, here 181.

[32] Barthes, “The Grain of the Voice” (cf. note 31), 182.

[33] Barthes, “The Grain of the Voice” (cf. note 31), 181.

[34] Dolar, *Voice and Nothing More* (cf. note 24), 22.

[35] Roland Barthes, “Music, Voice, Language,” in *Roland Barthes: The Responsibility of Forms: Critical Essays on Music, Art, and Representation*, trans. Richard Howard (Berkeley and Los Angeles: University of California Press, 1991 [1985]), 278–285, here 280.

[36] The term ‘technotrauma’ goes back to the international exploratory workshop *Techno-Trauma: From Analog to Digital* that took place at Humboldt University, Berlin, Department of Musicology and Media Studies, in April 2014.

[37] My translation. The German original reads: “Was ich fürchte, ist nicht die schreckliche Gestalt hinter meinem Stuhle, sondern ihre Stimme: auch nicht die Worte, sondern der schauderhaft unartikulierte und unmenschliche Ton jener Gestalt. Ja, wenn sie noch redete, wie Menschen reden!” Qtd. in Oliver Jahraus, “Heilige Texte,” in *Die Textualität der Kultur: Gegenstände, Methoden, Probleme der kultur- und literaturwissenschaftlichen Forschung*, eds. Christian Baier, Nina Benkert, and Hans-Joachim Schott (Bamberg: University of Bamberg Press, 2014), 39–56, here 42.

[38] Qtd. in Dolar, *A Voice and Nothing More* (cf. note 24), 7.

[39] Von Kempelen, *Mechanismus der menschlichen Sprache* (cf. note 22), 439.

[40] John Hollingshead, *My Lifetime.* Vol. I (London: Sampson Low, Marston & Company, 1895), 68; emphasis in the original.

[41] Hankins and Silverman, “Vox Mechanica” (cf. note 21), 215. Here with reference to “The Euphonia,” in *Illustrated London News* 9 (1846), 96.

[42] Cecil Thompson, “Artificial Voices Made in a Film Studio – Unspoken Words Heard from a Screen – Celluloid Marvel – An Englishman’s Eerie Invention,” in *The Daily Express London*, February 16, 1931, 1–2.

[43] For a more detailed description of Humphriss’ sound experiments, see Thomas Y. Levin, “‘Tones from Out of Nowhere’: Rudolph Pfenninger and the Archaeology of Synthetic Sound,” in *Grey Room* 12 (2003): 32–79.

[44] For a more detailed description, see Andrey Smirnov, “Synthesized Voices of the Revolutionary Utopia: Early Attempts to Synthesize Speaking and Singing Voice in Post-Revolutionary Russia (1920s),” in *Electrified Voices: Medial, Socio-Historical and Cultural Aspects of Voice Transfer*, eds. Dmitri Zakharine and Nils Meise (Göttingen: V & R Unipress, 2012), 163–185.

[45] Edgar Allan Poe, “The Facts in the Case of M. Valdemar,” accessed July 9, 2016, <http://www.eapoe.org/works/tales/vldmard.htm>; emphases in the original.

[46] Sigmund Freud, “The Uncanny,” in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, trans. James Strachey in collaboration with Anna Freud, Vol. XVII (1917–1919) (London: The Hogarth Press, 1955), 219–252, here 219.

[47] Ernst Jentsch, “Zur Psychologie des Unheimlichen,” in *Psychiatrisch-Neurologische Wochenschrift* 22 (August 1906), 195–198, here 197.

[48] Ernst Jentsch, “On the Psychology of the Uncanny (1906),” trans. Roy Sellars, last accessed July 9, 2016, <http://www.art3idea.psu.edu/locus/Jentsch_uncanny.pdf>.

[49] Masahiro Mori, “The Uncanny Valley,” trans. Karl F. MacDorman and Takashi Minato, in *Energy* 7.4 (1970), 33–35.

[50] For a programmatic example, see "The uncanny valley for computer speech": "What ever happened to the Stephen Hawking kind of speech device? It didn't creep you out because it was so clearly a machine." accessed July 9, 2016 <http://akinokure.blogspot.de/2014/01/the-uncanny-valley-for-computer-speech.html>.

[51] Mori, "The Uncanny Valley" (cf. note 49), 33–35.

[52] For a more in-depth analysis of Hatsune Mike and virtual idols, see Rafal Zaborowski, "Hatsune Miku and Japanese Virtual Idols," in *The Oxford Handbook of Music and Virtuality*, eds. Sheila Whiteley and Shara Rambarran (New York: Oxford University Press, 2016), 111–128.

[53] Vito Pinto also entitled a sub-chapter of his German-language PhD-thesis "Es gibt keine 'körperlosen Stimmen.'" Vito Pinto, *Stimmen auf der Spur: Zur technischen Realisierung der Stimme in Theater, Hörspiel und Film* (Bielefeld: Transcript, 2012), 182–185.

[54] Cf. Lynne Kendrick and David Roesner, introduction to *Theatre Noise: The Sound of Performance*, eds. Lynne Kendrick and David Roesner (Newcastle upon Tyne: Cambridge Scholars Publishing, 2011), xiv–xxxv, here: xiv.

[55] John Durham Peters's essay "The Voice and Modern Media" is a programmatic example of this assumption. On the one hand, he explicitly writes about disembodied voices with reference to the radio theorist Rudolf Arnheim (91–92); on the other hand, he argues that bodies imply voices and voices imply bodies (97) and that "Modern sound-recording media [...] are physiological: they magnify and amplify the body in its details. The voice figures bodies [...]" (98). John Durham Peters, "The Voice and Modern Media," in *Kunst-Stimmen*, eds. Doris Kolesch and Jenny Schrödl (Berlin: Theater der Zeit, 2004), 85–100.

[56] Peters, "The Voice and Modern Media" (cf. note 55), 215; my emphasis.

[57] Derrida, *Of Grammatology* (cf. note 1), 10.

[58] For the sake of completeness, it should be mentioned that about 30 years after *Of Grammatology* Derrida distinguished between speech and voice. In an interview from 2001, "Above All, No Journalists!", he argues that "the recording of the voice is one of the most important phenomena of the twentieth century" because "it gives living presence a possibility of 'being there' anew that is without equal and without precedent" (qtd. in Michael Naas, *Miracle and Machine: Jacques Derrida and the Two Sources of Religion, Science, and the Media* (New York: Fordham University Press, 2012), 142). This statement seems to be counterintuitive when one has his argumentation from 1967 in mind. Furthermore, in this interview he describes the phonographically primal scene of being affected by hearing the voices of dead people: "I am always overwhelmed when I hear the voice of someone who is dead, as I am not when I see a photograph or an image of the dead person. [...] I can also be touched, *presently*, by the recorded speech of someone who is dead. I can, *here and now*, be affected by a voice from beyond the grave"; emphases in the original (qtd. in Naas, *Miracle and Machine*, 143–144). Therefore, Derrida, too, was aware that the affective power of spoken language lies not in the code system of speech but in the individuality of voices, that is, the materiality of sound.

[59] Thomas Levin aims his project *Phonopost: Rediscovering a Forgotten Chapter of Media History* on this history of voice mail. The project is located partly at Freie Universität Berlin, funded by the Einstein Foundation. For more information, see <https://www.phono-post.org>.

[60] Frederick Garbit, *The Phonograph and its Inventor, Thomas Alva Edison* (Boston: Gunn, Bless & Co., 1878), 10–11.

[61] Cited in Barbara Engh, "Adorno and the Sirens: Tele-Phonographic Bodies," In *Embodied Voices: Representing Female Vocality in Western Culture*, eds. Leslie C. Dunn and Nancy A. Jones (Cambridge: Cambridge University Press, 1994), 120–135, here 122.

[62] Engh, "Adorno and the Sirens" (cf. note 61), 122; my emphasis.

[63] Marshall McLuhan, *Understanding Media: The Extensions of Man* (London/New York: Routledge 2009 [1963]), 291.

[64] Helen Blank, Alfred Anwander, and Katharina von Kriegstein, "Direct Structural Connections between Voice- and Face-Recognition Areas," in *The Journal of Neuroscience* 31.36 (2011), 12906–12915.

[65] Theodor W. Adorno, *Minima Moralia: Reflections from Damaged Life*, trans. E. F. N. Jephcott (London/New York: Verso, 1974), 111.

[66] Roland Barthes, "Music, Voice, Language" (cf. note 35), 280.