



'Yours in Speech (and Speakos)': Automatic Speech Recognition as Writing Technology in Richard Powers's Literary Practice

Lisa van Straten, Radboud University, The Netherlands, lisavanstraten@live.nl

Past decades have seen a rise in scholarly interest in the impact of digital technologies on literary production (see e.g. Berensmeyer 2022; Kirschenbaum 2016; Lyons and Marquilhaes 2017; Côrtes Maduro 2017). However, little attention has been paid to the influence of one such recent technological innovation: automatic speech recognition (ASR). This article explores the affordances and limitations presented by this modern-day writing technology through the case study of renowned author Richard Powers. It will discuss how speech recognition's hardware and software shaped Powers's writing process. Subsequently, it will critically evaluate Powers's claim that ASR facilitates a more immediate and authentic mode of storytelling by considering the technology's operation within a multifaceted network of remediation (Grusin and Bolter 2000). The second part of the article will complicate Powers's views further by analyzing his novel *Orfeo* (2014), which traces parallels between the author and its protagonist— a composer obsessed with wrestling free from the constraints of his medium in his quest for authentic artistic expression. Overall, this article will argue for the importance of understanding digital writing technologies as complex media networks that not only shape literary products but also their authors and modes of reception. Powers's embrace of dictation software thus not only unveils a fascinating chapter in the evolution of writing practices but also beckons a deeper exploration of how digital tools redefine the relationship between author, reader, and text.



On January 7, 2007, *The New York Times* published an article by Pulitzer Prize-winning novelist Richard Powers that caused quite a stir in literary circles. The article, titled 'How to Speak a Book', promised readers insight into how the renowned author crafted his lauded literary works. Powers's answer to this question, however, caught many off guard. Powers confessed to 'hav[ing] [rarely] touched a keyboard for years' and using automatic speech recognition (ASR) software to write his novels. In the article, Powers situates himself in a long line of authors who used forms of dictation as part of their writing process while delineating the new advantages and constraints of the advanced software. For him, working with this form of dictation provided a way of investing his writing with a greater sense of immediacy and authentic musicality than had been possible with earlier writing practices. This article will use Richard Powers as a case study to explore the affordances and limitations of speech recognition software as a writing technology and its impact on the relationship between author and text. It will do so on two levels: on the one hand, guided by Powers's own observations, the article will discuss how this modern technology informs the writing process and literary text through its advanced hardware and software. On the other hand, it will critically examine Powers's view that ASR facilitates a more immediate and authentic mode of writing. Concerning the latter, the article will turn to Powers's novel *Orfeo* (2014) to show how the book further complicates the author's earlier standpoint by shedding light on the entangled 'networks of remediation' that shape his work (Bolter and Grusin 2000: 64). As Powers states in his *The New York Times* article: 'Everything we write—through any medium—is lost in translation. But something new is always found again'. Therefore, in an attempt to broach new terrain in the study of digital writing technologies, this article aims to unpack what this translation process looks like for ASR, what new opportunities it presents and what limits it imposes upon artistic expression. As such, it seeks to illuminate not just how ASR reshapes the mechanics of writing but also how it brings into focus the frameworks within which artistic expression is conceived, mediated, and received.

Automatic Speech Recognition as Writing Technology

In 'How to Speak a Book', Powers begins the discussion of his dictation method by inviting the reader to witness him in the act of writing. He states: 'I write these words from bed, under the covers with my knees up, my head propped and my three-pound tablet PC [...] resting in my lap, almost forgettable. I speak untethered, without a headset, into the slate's microphone array. The words appear as fast as I can speak, or they wait out my long pauses' (2007a). The description challenges the conventional image of a writer at work; here, there is no pen or paper, not even a keyboard as a mechanical intermediary between author and text. Indeed, as Powers (2007a) points

out, the unnatural ‘finger mechanics’ required by ‘the 130-year-old qwerty keyboard’ no longer impede his cognitive flow as his ‘primary digital prosthetic doesn’t even have keys’. Instead, the ASR software allows him to speak freely, ‘unthethered’, as his speech is automatically translated into text without any effort on his part to convert his narrative into writing ‘one tedious letter at a time’ (Powers 2007a).

Powers (2007a) does not explain the actual workings of ASR technology in the article, except for brief references to microphones and his device’s ‘recognition engine’. As a result, readers get the impression that the technology is somewhat of a black box. Powers, however, was intimately acquainted with the software substructures of the machine. As a student at the University of Illinois, he became fascinated by the university’s pioneering system of educational computers known as Project PLATO. He dedicated his off-hours to learning how to program, buying his first computer in 1977 (Gross 2006; Dewey 2002: 7–8; Williams 2001: 97–98; Freeman 2006). In an interview with Jeffrey Williams (2001: 97), he reminisces about those ‘weekends [where] rather than going to a movie, [he] would go to the computer labs and look under the hood and try to figure out how they did what they did’. After his studies, he worked as a data processor and computer programmer in Boston before switching to a full-time writing career (Dewey 2002: 7–8; Williams 2001: 97–98; Freeman 2006; Vorda 2013). Despite this career move, digital technology remained one of his passions, ‘supply[ing] the bread and butter of [several] of [his] novels’ plots’ (Vorda 2013). As such, even though Powers (2007a) presents the technological medium as almost invisible and forgettable, it is fair to assume that he had at least a basic understanding—if not a comprehensive grasp—of its inner workings.

The nuts and bolts of ASR: capacities and constraints

Automatic speech recognition, or speech-to-text (STT), refers to the technology that enables computers to recognize and convert analog human speech into digital text.¹ This conversion process consists of roughly two stages, each marked by a specific software model. The first stage uses an acoustic model to translate the analog sound waves created by a person’s articulators (lips, tongue, epiglottis, and glottis) into digital data. During this stage, the vibrating air molecules that form the sonic waves of a speech signal are captured by a microphone and converted into a digital signal consisting of ones and zeroes. Software algorithms transform this digital input into a graph showing the signal’s amplitude (dB) over time. With the help of a Fast Fourier Transform algorithm, this information is used to create a more precise sound spectrogram, which visually represents the signal’s frequency (kHz) and intensity (dB) for overlapping

¹ The discussion of ASR in this section is based on Microsoft (2017), Mahmood (2017), and Van Lier (2018).

acoustic frames of 20–40 milliseconds. Since the 1980s, Hidden Markov Model (HMM) algorithms have been used to compare these frames to a dataset comprised of the acoustic features of phonemes. Phonemes are the smallest distinct sound units that make up a language and enable phonetic differentiation between words. As the sonic building blocks of a language, each phoneme consists of an acoustic signal with a specific frequency, intensity, and duration. HMMs, nowadays often combined with deep neural networks (DNNs), help indicate which acoustic frames correspond to what series of phonemes through statistical probability, thus converting the digital speech vectors into the most likely phoneme sequence.

In the second stage of ASR, a pronunciation model works to relate sets of phonemes to words. In this process, the acoustic features of phoneme sequences are compared to a pre-programmed phonetic dictionary constructed by linguists. ASR models created in the past two decades generally integrate this model into a statistical language model, which is used to determine the probability of word combinations through data-trained algorithms (also known as natural language processing (NLP)). Drawing on vast bodies of text, these algorithms learn to predict the likelihood of particular word pairings and patterns. While older ASR systems without NLP models can transform a speech signal into the most probable word sequence with considerable accuracy, NLP has further boosted its performance, especially when it comes to ASR's ability to disambiguate between similar-sounding words. At the end of this stage, the output of these models appears on a computer screen within a digital interface that enables users to read and edit the text.

These complex technological models provide ASR with unique affordances as a writing technology compared to earlier writing practices. In his article, Powers compares his method to earlier dictation-based practices of authors like Henry James, James Joyce, and Fyodor Dostoevsky. These authors narrated their stories to typists, allowing them to compose stories faster and more fluently. ASR, however, eradicated the need for such a human intermediary, enabling Powers to write even more rapidly with impressive accuracy. As Powers (2007a) observes: 'Whole phrases die and revive, as quickly as I could have hit backspace [...] It won't choke, even at bursts over 200 w.p.m.'. Released from pacing constraints, he 'can forget the machine is even there' and become fully immersed in his writing (Powers 2007a). He points out: 'I don't have to queue, stop, batch dispatch and queue up again. I spend less mental overhead on orthography and finger mechanics and more on hearing my characters speak themselves into existence' (Powers 2007a; see also Gross 2006). This immersion is further facilitated by his Motion LE1600 tablet's microphone array that allows him to move around, hide under the sheets, and turn off the lights, creating a sensory deprivation where he only experiences his own voice (Perell 2024: 1:17:14–1:18:40). By

eliminating the need for any physical act besides enunciation in the writing process, the software affords Powers ‘the freedom to be completely disembodied when [...] writing’ and reach what he considers a ‘pure compositional state’ (Michod 2007). The author can lose himself in his story, freed from some of the obvious constraints that faced his literary predecessors.

However, ASR’s greatest merit, according to Powers (2007a), lies in its foregrounding of language’s musicality—its primal dimension of meaning—that has been largely lost in an era ‘retooled for printed silence’. In an interview with Terry Gross (2006), Powers argues that people tend to subvocalize while reading because ‘much of the meaning of [a] passage comes about through the sound’. He considers subvocalization to be an evolutionary remnant of the times when storytelling was a primarily oral activity, one that shows ‘we still feel residual meaning in the wake of how things sound’ (Gross 2006). ASR technology relies heavily on this sonic dimension of language, as the software is programmed to understand the speech signal as an acoustic signal first and secondly as a lexical utterance. Powers only has to speak his words out loud rather than transform his utterances into letters on a page, making him more attuned to the acoustics of his story as opposed to its lexical form. In his own words, using ASR technology ‘allows [him] to think in terms of the music of the prose [...] [to] hear the rhythm and meaning of the passage [he’s] working on’ (Gross 2006). Doing so enables him to ‘shape a sentence so that it doesn’t just describe but actually participates in the emotion under description’ (John Hope 2016). Additionally, through ASR, he can experience his own ‘sentences and paragraphs as oral phenomena’, like his subvocalizing readers, attuning himself to a dimension of meaning hidden in their sonority rather than their denotation (Gross 2006). In this compositional state, his words seem to flow directly from his mind to the screen, almost magically, creating the illusion that there are no barriers between the story in his head and the one on the (digital) page (Michod 2007). Consequently, Powers (2007a) states, ASR technology invests his text with a never-before-seen sense of fluency, immediacy, and authenticity compared to earlier writing technologies. I will return to this point later in this discussion.

Although ASR offers unique affordances, the technology also presents unique challenges—the most apparent being accuracy. Even though, as Powers (2007a) points out, the software performs with great precision even at a high speech rate, its ability to make phonetic distinctions remains inferior to that of a human being. Humans instantly interpret speech within its specific context, whereas ASR does not understand the utterance as such but merely translates it by comparing its acoustic features to pre-programmed data sets. As a result, Powers (2007a) explains, these devices abound in ‘speakos and mondegreens’ that would never have occurred with a typist.

In his case, these appear even more frequently as the Motion LE1600 lacks a natural language processing model trained to model meaning based on textual probability and context (Powers 2007b). Because of this deficiency, '[his] tablet can turn [his] words hallucinatory without limit' (Powers 2007b). The end of his article illustrates that, as it includes two mishearings of his closing words: '[...] something new is always found again, in their eager years. In Derrida's fears. Make that: in the reader's ears' (Powers 2007a). When mistakes like this occur, Powers can tap on a word or phrase to respeak it (Powers 2007b). Whenever a word is not included in the software's pre-programmed dictionary, he can add it by handwriting with a stylus, navigating the on-screen keyboard, or spelling it out loud (Powers 2007b). Still, such inaccuracies inform the writing process to such a degree that Powers felt compelled to playfully end one of his online discussions of ASR technology with the phrase "Yours in speech and speakos" – a phrase that proved to be a fitting title for this article (Powers 2007b).

While these problems appear to be relatively minor hiccups in practice, they illustrate how ASR steers—and at times even requires—users to engage with it within certain input parameters to ensure a successful man-machine collaboration. For example, while Powers is free to wander around his room, the device's accuracy decreases rapidly when he speaks outside or in a noisy environment. As Powers (2007a) states: 'I've tried dictating to my tablet while rambling; traffic and birdsong make it babble'. Additionally, in order for the microphones to pick up his speech signal, he cannot move too far away from his tablet or speak indistinctly. The Motion LE1600 even required him to go through a whole program to 'train the system for [his] voice, speaking pace, and inflection' and 'define the directional angle that is to be used for speech recording' before use (Motion 2005: 56–57). Powers (2007a) observes how this set-up called for a 'huge cognitive readjustment' as he 'needed weeks to get over the oddness of auditioning [him]self in an empty room, to trust the flow of speech, to learn to hear [him]self think all over again'. In a way, one could say that while Powers was training the technology, so was the technology training him; through interacting with the device, Powers learned how to engage with the technology productively, modifying his input behavior in a cybernetic feedback loop.

Aside from matters related to position and environment, ASR's programming also places constraints on the contents of the speech signal. Software models, after all, are not neutral but 'programmed and engineered by [hidden] human actors' who invest the software with biases and a specific algorithmic agency (Klinger and Svensson 2018: 4654). As the 'author-initiator[s]' of these systems, these programmers decide their automated workings and limits, compelling users to interact with them on specific (if oftentimes hidden) terms (Hopfinger 2020: 171). Firstly, a user can only successfully

use ASR when speaking in a language that has been pre-installed on the device for speech recognition purposes. Using foreign vocabulary, slang, or newly invented words is discouraged, as such terms must be manually added to the device's pre-programmed dictionary. While adding a few words here and there presents only a minor interruption, making numerous corrections this way would greatly hinder the writing process. Consequently, the software is unsuitable as a writing technology for stories written in a dialect or minor language, multilingual literary works, or narratives including many neologisms. ASR programs based on NLP models are also ill-suited for authors writing experimental stories in which conventional modes of storytelling are upended and any logical connection between words and sentences is overturned. In these instances, the writing challenges the program's probability metrics, likely increasing its percentage of inaccuracies. Additionally, the software can prove problematic for authors writing in a secondary language, as accents can lead the technology to consistently misinterpret certain phonemes and words. Thus, even before Powers can attempt to write a story through ASR, both he and the story must meet these implied criteria in order to benefit from the technology's affordances. Put differently, even before a single word leaves his lips, Powers is programmed by the ASR technology to write within certain limits pertaining to speaking environment and volume as well as language, register, and style. Consequently, rather than functioning as a 'digital prosthetic' where the author has full control over the apparatus, the ASR technology creates a complex interdependence between author and device where the device influences the author as much as the other way around (Powers 2007a). As Maryla Hopfinger (2020: 55) points out in her study of digitally mediated literature: '[a] machine dictates its conditions, enforces the behaviour it determines, and makes humans dependent. While a brush, a pen, or a flute were tools controlled by a man who was able to subordinate them to himself/herself, all mechanical or electronic technologies gain power over people'. Thus, the technology determines what can be dictated by its user and how; the user can exercise control over the machine only within the specific action parameters set by the machine itself. From this angle, the technology seems hardly as 'forgettable' as Powers (2007a) suggests.

Immediacy, transparency, and mediation

Further examining Powers's claim that ASR restores a sense of unmediated immediacy and acoustic meaningfulness to writing through this lens reveals a far more complicated situation. In his article, Powers (2007a) implies that using ASR technology allows him to write more authentically than previous generations of authors, as his spoken narrative is transformed into digital text in real-time without any physical interaction with (hand)writing tools or interference from human intermediaries. The

apparent seamless nature of this process gives him the impression that the words on his tablet's screen are a rather direct and faithful representation of the story he composes within his mind. While Powers (2007a) briefly acknowledges the fact that there will always be a disconnect 'between the story in the mind and what hits the page' and that 'no interface will ever be clean or invisible enough' to achieve otherwise, his discussion of the technology's affordances suggests to the reader that ASR is, in fact, a step in that direction. What Powers disregards here, however, is that this modern writing technology is founded on and operates across many layers of (re)mediation that radically transform his words as they move from mind to mouth to page.

As Jay David Bolter and Richard Grusin (2000: 46, 48) have pointed out in their study of mediation, modern-day computers '[promise] the user an unmediated experience' by facilitating a smooth and instinctual user experience that makes the user forget the device is even there. In this strive for 'transparency', the technology tries to 'erase itself' and instill in the user 'a feeling that the medium has disappeared and the objects are present to him, a feeling that his experience is authentic' (Bolter and Grusin 2000: 70). From a technical point of view, this sense of 'transparency' relies on the computer's hidden 'layers of programming [that] always return control to the user', creating an illusion of autonomy despite the software's system of 'automated action[s]' (Bolter and Grusin 2000: 33). ASR is a great example of such an aspiring transparent digital medium, as the technology relies heavily on input from users, who remain largely unaware of the automated conversion processes initiated by their actions. Still, this sense of transparency does not mean that the transcribed text is therefore a(n) (more) unmediated form of expression, as Powers implies. In the first stage of ASR, for example, the expression transforms consecutively from a series of vibrating air molecules into a resonating diaphragm, then into an electrical current and eventually into a digital signal. Afterwards, this signal is broken into segments, converted into phonemes, and then combined and recombined into different words, sentences, and paragraphs using various algorithms. Finally, the digital output appears on the screen as text, which in itself is a remediation of the writing technology of pen and paper (Bolter and Grusin 2000: 45). The hardware and software thus rework Powers's supposedly authentic speech signal dozens of times, both in analog and digital ways, even before it takes the form of a transcription. Thus, the text on the screen is a highly mediated product of a network of remediation in which each 'act of mediation is dependent on another, indeed many other, acts of mediation' (Bolter and Grusin 2000: 56).

Moreover, even before his words undergo this process of technological remediation, Powers forms an integral part of this network. After all, the process begins with the remediation of his not-yet-fully-articulated thoughts, feelings, and memories into

coherent language. As Bolter and Grusin (2000: 57) have argued, language is not ‘a neutral, invisible conveyor of fully present meaning’ but ‘an active and visible mediator that fills up the space between signifying subjects and nature’ — i.e. a man-made system of phonologies, vocabularies, and morphosyntaxes that functions as a technology in its own right (see also Mufwene 2013; Hopfinger 2020: 166–167). Powers relies on these structures and their rules to construct the story in his mind. Therefore, even before it becomes a speech signal, Powers’s story is inherently shaped by the medium of language. Additionally, as Powers expresses his story, it is mediated by his body, specifically by the articulators discussed earlier. Here, a brain-muscle connection converts the story from silent thoughts into speech produced by different parts of the mouth, lips, and throat. Bruno Latour (1998) remarks on these interlocking ‘machines’ in a conversation with Powers: ‘Even when we are alone, in the flesh, we speak through all manner of machines: phonation, vocal chords, Broca’s area, English (for me a foreign artificial tongue, a huge distributed computer, for you a mother tongue a womblike web)’ (4). While I will not delve deeper into the mechanics of speaking, considering the scope of this article, these aspects are important to mention as they show how Powers himself functions as a mediating machine operating within a complex system of mediating technologies. Furthermore, it demonstrates that Powers’s presentation of the technology as a more authentic writing technology is based more on the device’s illusory transparency than reality. The advanced technology of ASR, in fact, adds even more layers of mediation to expression than earlier writing technologies, engaging its users in a dynamic media network where the intersecting affordances and limitations of language, the body, and technology dictate what can be written and how.

Technologies and Networks of Remediation in Powers’s Literary Writing

While Powers’s *The New York Times* article presents ASR as a liberating tool that enhances the immediacy and musical authenticity of storytelling, a closer look at the author’s body of work and interviews complicates this perspective. In many of his novels, such as *Galatea 2.2* (1995), *Plowing in the Dark* (2000), and *Playground* (2024), Powers thematizes the connection between technology, art, and society, exploring complex themes like the modern human experience, machine agency, and the nature of representation. These works reveal a nuanced understanding of how digital media shape human engagement with the world, be it through virtual reality, artificial intelligence or machine-driven applications (see Ghaleb 2020). Powers has long considered literature a powerful tool for conceptualizing and interrogating the man-machine entanglement of the modern age that followed ‘the transformation of the world through digitization’ (Freeman 2006; Williams 2001: 97). In an interview organized by the John Hope

Franklin Humanities Institute (2016: 26:57–27:15), he even stated ‘it seems to me [...] to tell the story of humans in the late twentieth century and early twenty-first century, if science and technology aren’t center stage, you’re not really getting it’. Part of the reason for literature’s appeal in this regard is, for him, the inherent intertwinement of literature and technology. In a conversation with Ines Ghaleb (2020: 298) he argues that ‘[a]rt is a technological endeavor, and every art possesses its own set of affordances, and enabling technologies’ (see also Perell 2024: 1:21:01–1:21:07). This perspective informs his understanding of the novel as a technological product, a ‘connection machine—the most complex artifact of networking that we’ve ever developed’ (Williams 2001: 105). He describes storytelling not as an inherently natural or primal activity, but as a structured system akin to a digital network that allows ‘free agents and the representations of agents [to] communicat[e] through narrow bandwidths across great distances’ (Latour 1998: 15). His terminology—‘machine’, ‘networking’, ‘bandwidths’—emphasizes his view of the novel as a form of digital communication, aligning writing more closely with software programming than traditional notions of literary authorship. Powers even draws direct parallels between fiction writing and coding. According to him, his background in programming ‘gave [him] many ways of thinking about form and structure as a fiction writer’ (Freeman 2006) and even inspired him to write an ‘aesthetic and narratological look at computer coding’ sometime in the future (Michod 2007). From this angle, Powers’s use of ASR can be understood not just as a pragmatic writing choice but as an integral part of his ongoing exploration of how digital technologies shape modern-day existence, including artistic expression. His literary works are not just about technology; they are intrinsically shaped by the technologies and man-machine interrelation thematized in his writing.

Indeed, his fascination with digital technology prompted Powers to experiment with ASR software in its early stages, eventually adopting it as his primary writing method for *The Echo Maker* (2006) (Powers 2007b). ASR remained central to his composition process for subsequent novels, including *Generosity* (2009) and *Orfeo* (2014). Over time, however, Powers began to use the technology more sparingly, incorporating it selectively alongside other writing media based on his mood and the specific demands of each book and scene (SRF 2021: 4:29–4:58). As he explains in an interview with David Perell (2024: 1:20:26–1:21:26):

‘[T]hink of it the way that a musician would use different instruments. [...] [I]f you’re writing a song you might reach for a guitar for a certain kind of song. You might sit down at the piano for a different kind of song [...] Or you might rotate. You might try different combinations in sequence or in parallel. I think the same goes for writing. I

think the tools that we use to write, we reach for them when we need them, when we detect through our intuition or through our intellect that we need to slow down or be more quiet or speed up, be more lively. And each tool has its affordances and allows you to get to different places’.

Through this analogy, Powers foregrounds how different writing technologies generate distinct literary effects as each medium inherently shapes the writing in unique ways. His choice of a musician for the analogy is particularly significant within the context of ASR: dictation made him more attuned to the primal musicality of language, leading him to tell his stories with a greater focus on their acoustics. Because of the affordances of ASR, Powers began using language as a kind of music, assuming the role of a bard-like composer rather than that of a traditional writer. This view of artistic expression as a technologically mediated act will prove especially revealing in *Orfeo*, where the novel itself becomes a site for showcasing and interrogating the expressive and structural implications of (digital) media for musical art.

Listening to Orfeo

The analogy between writer and musician is central to the construction of *Orfeo*, the story of an ambitious composer named Peter Els who, like a modern-day Orpheus, devotes his life to creating radically innovative music. In the following section, I will demonstrate how Powers uses Els’s pursuit of authentic musical expression to thematize the relationship between artist and (digital) media—one that closely parallels his own. *Orfeo*, as I will demonstrate, maps the complex networks of remediation that shape Els’s work, ultimately revealing that direct, transparent artistic expression is merely an illusion. As such, *Orfeo* serves as the literary counterpart to Powers’s *The New York Times* article, one that foregrounds the layers of (re)mediation between artist and (digital) technology that remain largely unexamined in his journalistic account.

In *Orfeo*, Powers presents Peter Els’s lifelong struggle with the medium of music in a way that resembles his own engagement with language and ASR. From a young age, Els is fascinated by sound, having grown up in a musical household with a piano-playing mother and music-obsessed father (14–17; Vorda 2013). As an eight-year-old boy, he starts playing the clarinet, and by age twelve, he spends his evenings immersed in his father’s extensive music collection, familiarizing himself with the classical compositions of Bach, Mozart, and Beethoven (16–17; Vorda 2013). His artistic ambitions take shape when he meets Clara, his first love, who introduces him to the great musical experimentalists of the twentieth century (31–32). Inspired to become a composer himself, Els studies music in college, focusing on post-war avant-garde composers

like Olivier Messiaen, Dmitri Shostakovich, and John Cage. He becomes obsessed with revolutionizing the medium and pushing the boundaries of what music can be. As the narrator of *Orfeo* observes, Els ‘staked his life on finding that larger thing. Something magnificent and enduring [that] hid under music’s exhausted surface’ (10–11). This artistic quest, however, ultimately proves unsuccessful, leading him to turn to DNA as a medium for musical experimentation. DNA allows him to create the masterpieces he always dreamed of, but at great cost to himself and his art. Els’s life story and love for music parallel Powers’s own in many ways (see Vorda 2013; McCracken 2014; John Hope 2016). However, more importantly, Els’s struggle with his medium reflects Powers’s own evolving relationship with ASR: Powers also was driven by a similar hope of uncovering new creative freedoms in digital technology, leading him to gain a deeper understanding of these systems’ built-in biases, affordances, and limitations. Indeed, many of Powers’s observations regarding the role of (digital) media in literary expression discussed previously can be seen reflected in *Orfeo*, transposed to the context of musical composition.

The mechanics of music

The connection between literature and music is made explicit throughout the narrative, as Powers draws implicit and explicit parallels between his own work as a writer and Els’s musical compositions.² Music is repeatedly described in *Orfeo* as a ‘language’ (100) that—though often wordless—communicates with an audience much like a novel through ‘musical idiom[s] [...] [and] [a] harmonic vocabulary’ (292). Over the course of his life, Els comes to believe that music has turned into an overdetermined language and has become exhausted as an artistic medium. He laments that music has lost its ability to serve as a ‘direct transcription of inner states’, feeling it has become ‘bogged down by meaning’ (329). The reason for that, Els observes, is the body, which is programmed to assign meaning to acoustic signals (329–330). Rather than experiencing music as mere sound, the body instinctively transforms music into a story: ‘The body had evolved to feel fear, hope, thrill, and peace in the presence of certain semi-ordered vibrations’ (330). Powers explains this further in an interview with Keenan McCracken (2014), where he describes himself as ‘the kind of listener who experiences music, if not as a language, then at least as an information-rich linguistic phenomenon’. Interweaving music and literature in his own writing, he states, feels

² The connection between language and music is also foregrounded in the novel through its structure, which appears to resemble a fugue (see McCracken 2014) or a ritornello (Leonard 2014), and the many detailed descriptions of musical compositions in which music is remediated through language (see Reichel 2017). For an in-depth discussion of the novel as a musical composition inspired by the Orfeo myth, see Miriam Fernandez-Santiago (2019) Kathryn Hume (2017) and A. Elizabeth Reichel (2017).

natural to him as '[his] love of harmonic expectation and progression in music is not unlike [his] love of cadence and register in prose' (McCracken 2014; see also Perell 2024: 44:38–44:43). Over centuries, musical traditions reinforced this tendency to connect music with linguistic meaning, embedding emotions and even storylines into particular combinations of notes, harmonies, and melodies. Thus, music gradually evolved from meaningless sounds into its own semiotic system.

While twentieth-century avant-garde composers sought to break with these conventions, Els discovers that even radical musical experimentation eventually solidifies into familiar patterns (326). This realization is echoed by Powers, who observes: 'the problem with the mind's desire to understand and embrace is that we ultimately kill the joy with overfamiliarity' (McCracken 2014). Els saw himself as standing at the 'moment of mash-up' following '[t]en short centuries [that] had burned through all available innovations, each more fleeting than the last' (30). As part of this new generation tasked with broaching new musical terrain, he dedicates his life to searching for '[f]resh, surprising music that escaped all human conventions' (73); a type of artistic expression that relies on pure sound that would transcend the body's built-in interpretational structures (see also Hume 2017). Els shares this goal with Powers, who likewise turned to sound, convinced that the medium would convey his internal state transparently. Yet, both author and protagonist initially neglect to explore how their media rely on a network of mediating technologies—both analog and digital—that fundamentally shape the act of artistic creation and their art. In composing his musical works, Els is no freer in bending music to his will than Powers is in crafting his narrative with ASR software. In both cases, the dream of unmediated expression is ultimately confronted by the reality of media constraints.

Indeed, Els soon discovers that music, far from being a transparent medium, is itself a technology, a structured system that both enables and restricts artistic creation. Music, after all, like language, is an artificial construct: it consists of '[t]welve chromatic pitches' (354), which, comparable to phonemes, serve as its fundamental building blocks. It also relies on notation, a system that, while preserving musical ideas, simultaneously bogs down pure sound, much like written language. As Els puts it: '[t]hat was the curse of literacy: Once you started writing music down, the game was half over' (29). The twelve chromatic pitches form the musician's 'toolkit' (100), making up 'the palette for everything from sultry seduction to funeral mass' (82) as if they were the 'meshed gears' of a 'magnificent timepiece' (18). While these structures inherently inform music, shaping its existence, Els begins to consider them 'twelve repeating black and white prison bars' (368) that form a 'straitjacket' (81) and 'harmonic jail' (20) he cannot escape: no music exists outside of these twelve pitches and no music can ever be written that transcends them. Any musical composition will

always necessarily imitate others, each being just one more variation on the same set of pitches. As Els points out: ‘A hundred thousand years of theme and variations, every composer stealing from every other’ (332). Linking this back to ASR, dictation software, at its core, is built on this principle of imitation: it is programmed to understand any utterance as a variation on the same set of phonemes, which form the foundation of the machine’s speech recognition model. As indicated before, the software even encourages its users to express themselves in a conventional or derivative manner to maintain a high level of accuracy. ASR’s programming thus reveals language to be a phonemic prison much like music. Even though variation in literature and music seems endless in practice, the finitude of these media themselves makes the number of possible variations theoretically finite as well. Consequently, as Els comes to realize, innovation is only possible within the limits set by the medium itself.

Furthermore, Els gradually recognizes that music does not exist in isolation but within an inescapable network of remediation in which different analog and digital technologies mediate the artistic process. All of these intersecting affordances and limitations dictate what Els can or cannot compose. As A. Elisabeth Reichel (2017: 81) has observed, music has become ‘encode[d] by means of historically contingent, socially and culturally constructed rules’, many of which depended on its supporting ‘materials and technologies’. For centuries, composers were restricted by the musical abilities of singers, instruments, and musicians, but over time, new tools redefined the contours of musical possibility. Musical production, for example, evolved radically over the past century with the advent of various recording and audio devices. *Orfeo* frequently foregrounds these technologies, referencing LPs, speakers, headphones, and MP3 players (e.g. 6, 32, 40, 72, 74, 86, 89).³ In the digital age, however, Els can wholly transcend many of the limitations imposed on his predecessors with the help of ‘a computer [programmed] to generate a string quintet using probability functions and Markov chains’—the same algorithmic principles that underlie ASR (72). He describes how ‘[t]he computer made it possible to shape any pitch, amplitude, timbre, and duration’, allowing musical fragments to be ‘combined and recombined, slowed, sped up, inverted, reversed, stacked into evolving rhythms and incanted in banks of antiphony’ (97). He witnesses how, over time, these digital tools transformed into Sibelius, a software model sophisticated enough to ‘[turn] an average tunesmith into Orpheus’ with ‘cut-and-paste harmonies, point-and-click tone painting, one-button transposition [...] [that] [transform] a handful of raw building blocks [into] a new two-minute stunning tutti’ (319). The model eliminates many of the practical difficulties and

³ Powers also draws attention to the novel’s remediation of digital technologies by inserting Els’s Twitter messages, marked by distinctive formatting, into the main narrative. In this way, the reader experiences digital text through analog text.

extensive labor involved in composing and creating scores through analog means. Yet Els finds this freedom unsatisfying. He considers the software's affordances in terms of speed, efficiency, and convenience to be incommensurate with the loss of traditional tools and the artist's involvement in the composition process, leading him to '[yearn] for the clumsy, freighted flights of earthly instruments' (97). Sibelius much resembles a text-generating machine as it combines and reworks pre-programmed musical phrases with minimal input from the artist, obscuring the boundary between machine-made and man-made. Els's stance on Sibelius thus raises complex questions regarding the role of the artist in semi-automated digital creation processes that also pertain to ASR. For example, at what point does the artist lose (primary) authorship over his/her work? And is art's worth dependent on the lingering presence of the artist's hand and the limits of his/her abilities and materials?

Switching media: biological melodies

While the novel does not attempt to formulate an answer to these questions, it does provide a compelling thought experiment in Els's use of DNA for musical composition. From early on in his life, Els considered music's 'pattern language' (30) to be connected to other systems for representing the world, amongst which the mathematical language of chemistry: 'To [him], music and chemistry were each other's long-lost twins [...] The structures of long polymers reminded him of Webern variations. The outlandish probability fields of atomic orbitals [...] felt like the units of an avant-garde notation' (57). Powers echoed this observation, arguing that '[b]oth science and art consist of endless propagating experiments and speculations of what is and what might be'; two 'languages' that merely use different conversion keys to map the world (McCracken 2014). As Els got older, experimental composers sought to blur the line between the two disciplines even more by making 'music from everything. Fugues from fractals. A prelude extracted from the digits of pi. Sonatas written by the solar wind' (142). Additionally, at this time, a school of musical thought emerged around 'biocomposing'; music based on mathematical patterns found within the body (330). As Els observes: 'Brain waves, skin conductivity, and heartbeats: anything could generate surprise melodies' (330). One night, he hears compositions based on DNA sequences on the radio, which inspire him to compose his own music with biological material: 'soundtracks extracted from DNA—strange murmurings transposed from the notorious four-letter alphabet of nucleotides into the twelve pitches of the chromatic scale. But the real art would be to reverse the process, to inscribe a piece for safekeeping into the genetic material of a bacterium' (333). DNA appears to Els a 'virgin territory', 'a newfound land' free of 'the twelve black and white bars in front of musical freedom', as the medium is made up of patterns that are strange to the human ear, resisting any conventional cognitive-

aural framework of understanding (334, 355). Put differently, he believes the genetic material will provide him with '[a] grammar but no dictionary, sense but no meaning', a musical language outside of music's overdetermined semiotic system (213).

Changing his medium, however, does not free him from constraints on his artistic expression. Rather, in writing music through DNA, he participates in a whole new network of remediation that radically undermines the very essence of music itself. Despite being a virgin medium for composing, DNA comes with its own set of limitations, requiring several technologies and layers of mediation. For example, in order to transform the genetic material into a 'jukebox', Els has to convert music 'into a string of zeros and ones, [which] [are] then converted again into base four' (359). Subsequently, he needs to '[divide] those two numbers to produce a short key' that is embedded in the material with the help of advanced laboratory tools as if he were 'put[ting] the tape inside the player' (359, 333–334). Thus, he transforms analog music into a digital signal and back into an analog format, after which the musical sequences in the cell mutate, creating an 'endless change in the musical message' (334). Els considers this continual change to be 'more like a feature than a bug', thus underlining his perception of DNA as a software algorithm. He transforms himself into a 'biohacker', a composer-turned-programmer who writes a musical code within the genetic material that will, in turn, keep rewriting its own script until eternity (356). As such, he trades the role of author for that of an 'author-initiator' to borrow Hopfinger's term (2020: 171). In fact, he has no other choice, since the limitations of his own bodily apparatus—his hands, his mind, his ears—limit his own power over the genetic medium. His creations cannot be seen, heard, or played, having become almost unrecognizable through the medial constraints of the DNA. Thus, the constraints of the medium and his own human capabilities inherently shape his artistic expression in new ways, further undermining Els's dream of authentic artistic expression. Els, however, is drunk on the promise of a musical revolution, exclaiming: 'What's more beautiful than music you can't hear?' Still, his words leave the reader wondering: what's the point of music you can't hear?

In composing music through the automated self-replicating chains of DNA, Els brings his 'merely human capacities to the sphere of the transhuman-superhuman bordering the divine' (Fernandez-Santiago 2019: 135). But in the process, he loses music's essence, its defining characteristic: its ability to move audiences through sound. In Els's pursuit, superhuman artistic transcendence comes at the cost of a human recipient. While Powers does not actively frame this development in one way or another in the novel, the reader cannot but feel that somehow such a bargain defeats art's purpose. Consequently, the narrative appears to raise the question: is art's ability to be experienced ultimately more important than radical medial innovation? In a world where artists are continually transcending the limitations imposed on them by

their bodies and minds through digital technologies, are the experiential boundaries of artist and audience still important (enough) to cater to? In an interview, Powers seems to suggest so, observing that ‘if you throw away everything of significance in the name of the new and in the name of attention and transformation, then you’re left asking people to attend to something that isn’t human, that is too alien to give any kind of solace or understanding’ (McCracken 2014). Thus, he argues that conventions and constraints in art are useful as they allow an artwork to communicate with an audience—take away all that is familiar and an audience will have no way of relating to the work. Thus, while the structural limitations imposed on the artist appear to stand in the way of innovation, he observes, they perform the vital task of anchoring the work’s connection to the audience.

Els’s DNA compositions present a far more extreme vision of digital technology’s impact on artistic creation than ASR, pushing the idea of machine-assisted authorship to a radical degree. While ASR modifies the writing process by automating transcription and, increasingly, the generation of text, it remains tied to human cognition: writers still shape their ideas, revise their sentences, and impose their own creative vision on its in- and output. By contrast, Els’s turn to DNA as a compositional medium imagines a form of artistic creation that entirely transcends human sensory experience and interpretive control. His compositions are encoded into biological material that evolves independently, making them not just difficult to access but fundamentally unstable, shifting beyond the artist’s grasp in ways that no ASR-generated output would. In this way, *Orfeo* envisions a future in which digital mediation does not simply assist or enhance human creativity but ultimately takes over artistic control altogether, out of human reach. Additionally, if ASR raises concerns about diminished authorial agency due to the influence of software programming, Els’s work opens up an even more pressing question: what happens when art ceases to be made *for* human perception at all? His DNA music is not just unheardable and unplayable by traditional means—it is conceptually alien, existing in a realm where neither the artist nor the audience can fully comprehend or engage with it. While ASR still operates within the boundaries of human language and cognition, Els’s compositions gesture toward a post-human form of art that no longer relies on human senses, intuition, or meaning-making. The novel thus sketches an image of a digital future in which art may not merely involve new tools that expand human creativity but could lead to artistic production that is entirely untethered from human experience itself. Simultaneously, however, *Orfeo* reminds us that art relies on meaning generated within a network of remediation that includes not just writing technologies but also authors and readers. Thus, while future literary endeavors involving digital media may revolutionize storytelling, giving rise to new writing technologies and experimental narratives, it is important to keep an eye out for its

impact beyond that; how expression is restricted by new networks of remediation, how authorial agency is shared or even transferred to technologies, and how, potentially, the price of progress will be the presence of a beholder. Literature, as Powers's work testifies to, plays a pivotal role in exploring the complexities and (philosophical) impact of these networks as they evolve to include increasingly advanced digital technologies. In tracing these speculative frontiers, literature becomes a kind of interface itself—a place where we reckon with the promise and peril of art no longer made with or accessible to us.

Conclusion

Powers's embrace of ASR signals a fascinating new chapter in the history of writing technologies, opening up novel ways of thinking about the interconnection between literature, digital technologies, and artistic expression. As this article has shown, even though digital media promote an illusion of transparency, their often hidden programming shapes the author's writing practice in tangible and intangible ways. Consequently, despite its seeming facilitation of a more authentic and immediate writing process, ASR operates within a complex network of remediation that interrelates author and machine. In this process, literature becomes a product of digital co-creation, upending traditional notions of authorship and signaling a new era of literary creation in which the artist can expand and even transcend his/her own abilities. Particularly in his novel *Orfeo*, Powers presents a striking meditation on the entanglement of art and technology, exploring how creative expression is always shaped by the media networks within which it materializes. Whether through the algorithmic composition of Sibelius or the radical experiment of encoding music into DNA, the novel repeatedly asks: how does technological mediation alter not only the act of creation but also the very nature of what can be created? As such, it can be said to function as a literary product as well as a self-conscious dramatization of Powers's interest in human-technology interaction. Through Els's journey, Powers reveals both the intoxicating potential and the artistic cost of transcending human limitations through technology. While digital tools promise new frontiers of artistic possibility, they also threaten to sever art from its defining essence: its capacity to communicate, to be *heard*. In his yearning for absolute freedom, Els ultimately confronts the paradox that to escape artistic constraints is to undermine art's very existence. And yet, even as *Orfeo* chronicles the dissolution of authorship into networks, codes, and self-perpetuating algorithms, it also suggests that true artistic expression persists—not in spite of constraints, but because of them. Like a melody struggling against silence, or a voice shaping itself against the blankness of a page, art finds its form within the very limits that seem to confine it, travelling from one mind to another—imperfect, remediated, and limited, but powerful nonetheless.

Competing Interests

The author declares that she has no competing interests.

References

- Berensmeyer, Ingo. 2022. *A Short Media History of English Literature*. De Gruyter. <https://doi.org/10.1515/9783110784459>.
- Bolter, Jay David, and Richard Grusin. 2000. *Remediation: Understanding New Media*. MIT Press.
- Côrtés Maduro, Daniela, editor. 2017. *Digital Media and Textuality: From Creation to Archiving*. transcript Verlag.
- Dewey, Joseph. 2002. *Understanding Richard Powers*. University of South Carolina Press.
- Fernandez-Santiago, Miriam. 2019. "Of Language and Music: A Neo-Baroque, Environmental Approach to the Human, Infrahuman and Superhuman in Richard Powers's *Orfeo*." *Anglia* 137(1): 126–146. <https://doi.org/10.1515/ang-2019-0008>.
- Freeman, John. 2006. "Richard Powers: Confessions of a Geek." *Independent*, Dec. 15. <http://www.independent.co.uk/arts-entertainment/books/features/richard-powers-confessions-of-a-geek-428433.html>.
- Ghalleb, Ines. 2020. "The Interdisciplinary Mind: Modes of Evolution in Richard Powers's *Galatea 2.2*, *Plowing in the Dark*, *The Echo Maker*, and *Generosity: An Enhancement*." PhD diss., Ludwig-Maximilians-Universität. http://www.edoc.ub.uni-muenchen.de/27787/1/Ghalleb_Ines.pdf.
- Gross, Terry. 2006. "Author Richard Powers." *FreshAir Podcast*, Dec. 12. <https://freshairarchive.org/segments/author-richard-powers>.
- Hopfinger, Maryla. 2020. *Literature and Media: After 1989*. Translated by Andrzej Wojtasik. Peter Lang.
- Hume, K. 2017. "Novelty, Pattern, and Force in Richard Powers's *Orfeo*." *Orbit: A Journal of American Literature* 5(1): 1–19. <https://doi.org/10.16995/orbit.202>.
- John Hope Franklin Humanities Institute. 2016. "Words and Music: A Conversation with Richard Powers." YouTube, Oct. 12. www.youtube.com/watch?v=ec5MCA0e6Og&t=912s&ab_chan=nel=DukeFranklinHumanitiesInstitute.
- Kirschenbaum, Matthew G. 2016. *Track Changes: A Literary History of Word Processing*. Harvard University Press. <https://doi.org/10.4159/9780674969469>.
- Klinger, Ulrike, and Jakob Svensson. 2018. "The End of Media Logics? On Algorithms and Agency." *New Media & Society* 20(12): 4653–4670. <https://doi.org/10.1177/1461444818779750>.
- Latour, Bruno. 1998. "Two Writers Facing One Turing Test—A Dialog in Honor of HAL Between Richard Powers and Bruno Latour." *Common Knowledge* 7(1): 1–17. www.bruno-latour.fr/sites/default/files/P-72%20POWERS-DIALOG-2-97.pdf.
- Leonard, Andrew. 2014. "The astonishing power of Richard Powers." *Salon*, Feb. 9. www.salon.com/2014/02/09/the_astonishing_power_of_richard_powers/.
- Lyons, Martyn, and Rita Marquilhas, editors. 2017. *Approaches to the History of Written Culture: A World Inscribed*. Palgrave Macmillan.

- Mahmood, Zafar. 2017. "Lecture 9—Speech Recognition (ASR)." YouTube, March 15. www.youtube.com/watch?v=HyUtT_z-cms&t=2256s&ab_channel=ZafarMahmood.
- McCracken, Keenan. 2014. "A Conversation with Richard Powers." *Music & Literature*, Sept. 11. www.musicandliterature.org/features/2014/9/4/a-conversation-with-richard-powers.
- Michod, Alec. 2007. "An Interview with Richard Powers." *The Believer*, Feb. 1. www.thebeliever.net/an-interview-with-richard-powers/.
- Microsoft Research. 2017. "Automatic Speech Recognition—An Overview." YouTube, Sept. 11. www.youtube.com/watch?v=q67z7PTGRI8&t=3227s&ab_channel=MicrosoftResearch.
- Motion Computing. 2005. "LE-Series and LS-Series Tablet PCs, Microsoft Windows XP Tablet PC Edition 2005: User Guide." *Motion Computing Inc.* www.manualpdf.in/motion/le1600/manual?p=2.
- Mufwene, Salikoko. 2013. "Language as Technology: Some Questions that Evolutionary Linguistics Should Address." In *In Search of Universal Grammar: From Old Norse to Zoque*, edited by Terje Lohndal. John Benjamins Publishing Company. <https://doi.org/10.1075/la.202.22>.
- Perell, David. 2024. "Meet Pulitzer Prize-Winning Stanford Professor—Richard Powers." YouTube, Oct. 9. www.youtube.com/watch?v=QUDlpMN-f5w&ab_channel=DavidPerell.
- Powers, Richard. 1995. *Galatea 2.2*. Harper Perennial.
- Powers, Richard. 2000. *Plowing in the Dark*. Farrar, Straus and Giroux.
- Powers, Richard. 2007b. Post to "Echoes and real voices." *Ascent Stage*, Jan. 17, 12:02 pm. www.ascentstage.com/archives/2007/01/echoes_and_real-html/.
- Powers, Richard. 2007a. "How to Speak a Book." *The New York Times*, Jan. 7. www.nytimes.com/2007/01/07/books/review/Powers2.t.html.
- Powers, Richard. [2014] 2019. *Orfeo*. Atlantic Books.
- Powers, Richard. 2024. *Playground*. W. W. Norton & Company.
- Reichel, Elisabeth A. 2017. "Musical Macrostructures in *The Gold Bug Variations* and *Orfeo* by Richard Powers; or, Toward a Media-Conscious Audionarratology." *Partial Answers: Journal of Literature and the History of Ideas* 15(1): 81–98. <https://doi.org/10.1353/pan.2017.0005>.
- SRF Kultur Sternstunden. 2021. "Richard Powers über die Entfremdung des Menschen von der Natur." YouTube, Dec. 7. www.youtube.com/watch?v=f_z3-h7c-Ck&t=286s&ab_channel=SRFKulturSternstunden.
- Van Lier, Hannes. 2018. "A Basic Introduction to Speech Recognition (Hidden Markov Models & Neural Networks)." YouTube, Sept. 3. www.youtube.com/watch?v=UOXtE4_QLXI&ab_channel=HannesvanLier.
- Vorda, Allan. 2013. "A Fugitive Language: An Interview with Richard Powers." *Rain Taxi*. www.raintaxi.com/a-fugitive-language-an-interview-with-richard-powers/.
- Williams, Jeffrey. 2001. "The Last Generalist: An Interview with Richard Powers." *Minnesota Review* 52(54): 95–114. www.muse.jhu.edu/article/438876.

