digitale gesellschaft | NRW

Marcus Erbe / Aycha Riffi / Wolfgang Zielinski (Hrsg.)

Mediale Stimmentwürfe

Perspectives of Media Voice Designs

Schriftenreihe zur digitalen Gesellschaft NRW

7

kopaed

Marcus Erbe / Aycha Riffi / Wolfgang Zielinski (Hrsg.)

Mediale Stimmentwürfe

Perspectives of Media Voice Designs

Düsseldorf – München
www.kopaed.de

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

ISBN 978-3-96848-642-0

In der Schriftenreihe zur digitalen Gesellschaft NRW vertreten die Autorinnen und Autoren ihre eigene Meinung, ohne dass diese notwendigerweise die Meinung des Landes Nordrhein-Westfalen widerspiegelt.

Die Veröffentlichung entstand mit freundlicher Unterstützung der Staatskanzlei des Landes Nordrhein-Westfalen.

Verlag: kopaed verlagsgmbh
Umschlaggestaltung: Georg Jorczyk

Grimme-Institut – Gesellschaft für Medien, Bildung und Kultur mbH, Marl 2022

Eine Open Access Version dieses Bands ist zu finden unter:
http://www.grimme-institut.de/schriftenreihe

Lílian Campesato, Fernando Iazzetta

Voice as a Resonance of Listening

Abstract: In diesem Beitrag wollen wir bildliche Darstellungen der Stimme in verschiedenen Kontexten untersuchen. Ausgehend von der Resonanz im Sinne eines Elements, welches sowohl das repräsentiert, was den Klang erzeugt, als auch das, was vom Klang beeinflusst wird, werden wir einige Beispiele dieses Prozesses der Bilderzeugung durch Stimme im künstlerischen, technologischen und kommunikativen Kontext diskutieren.

Abstract: In this article, we are interested in exploring the voice's imagetic representations in different contexts. Starting from the idea of resonance as an element that represents both what produces sound and what is affected by sound, we will address some examples of this process of image production through voice in the artistic, technological and communicational contexts.

1 Voice and Listening

Vibration is a condition of what is in motion, of what is alive. It is the movement of a mechanical body in relation to a state of equilibrium. Everything that moves acts on its surroundings, creating frictions, resonances, shocks, impulses. Hence the importance of vibration for us to exist: It is an index of our existence, as well as of our relationship with everything that permeates our surroundings. Like most living beings, we learn to vibrate with the world as a way of interacting with it. And being mechanical, the vibration invokes the action of, and on, our bodies: It is an indication of presence. Our senses can perceive vibration in two dimensions. One, relatively slow, involving the movement of large bodies that make us tremble. Another, very quick and subtle, is unable to activate the tactile sensors spread throughout our bodies, but manifests itself in the form of what we call sound.

Sound is a very fast, but very weak, pressure variation that propagates through some material medium. Our bodies are equipped with a very precise system for both producing and perceiving sounds. Although phonatory and auditory mechanisms are usually treated as independent, autonomous devices, they act in a complementary way in our interaction with the vibrational world. Using one's voice and listening are

interdependent forms of action and there is no way to fully understand one without considering the other. This close connection between voice and listening can be approached in several ways. It implies the correspondence between the speech and hearing apparatuses, the interrelationship between individuals in communication processes, and the various forms of interaction between bodies and the environment.

In this article, we are interested in exploring the imagetic representations of voice in different contexts and exploring its connection with hearing. Starting from the idea of resonance as an element that represents both what produces sound and what is affected by sound, we will address three examples of this image-making process through voice in artistic, technological and cultural contexts: a 'silent' sound work by Christian Marclay; the invention of virtual popstar Hatsune Miku; and the unusual record of a 'mythical' character in Brazilian popular culture.

We take voice as a relational and multimodal process to emphasize its strong connection with the physical, corporal, and material domains. A voice represents a body, a space, and evokes a series of poetic, symbolic and affective instances. It is a trait, a mark, a sign. A voice is always someone's voice, or the voice of something. As Don Ihde points out, "all sounds are in a broad sense 'voices', the voices of things, of others, of the gods, and of myself" (2012, p. 147). On the other hand, voice is a powerful source of images:[1] It has the capacity to represent everything – histories, memories, bodies, spaces and times – that resonates when it sounds. However, this whole process always depends on listening, because without listening, the voice is mute.

By the middle of the last century, Alfred Tomatis, a French physician who dedicated himself to studying the relationships between voice, hearing and cognitive processes, speculated about the feedback chain involving voice and listening: "the voice only reproduces what the ear hears" (Tomatis 1992, p. 49). Analyzing some singers' disorders that prevented them from emitting certain sounds, Tomatis studied their auditory curves and realized that they "showed a hearing loss at the same frequency level" (Tomatis 1992, p. 41). Based on empirical inferences, not always demonstrated with scientific support,[2] Tomatis perceived a causal relationship between the composition of the listening and vocal spectra, leading him to affirm that "a person can only reproduce vocally what he is capable of hearing" (Tomatis 1992, p. 44). Or, stated in another way: "One sings with one's ear" (ibid). Controversial in his methods and treated with reservations by his peers, Tomatis' work had received enough recognition to attract the attention of renowned artists such as Maria Callas, Sting and Gérard Depardieu, who sought him out to solve their vocal shortcomings.

The connection between voice and listening envisioned by Tomatis should not be overlooked. Particularly when considered from an evolutionary point of view, it is assumed that voice and listening have developed in an integrated manner, at least for vertebrates, due to the relevance of sound communication among members of the same species. As Carl Gans suggests, when individuals acquire the ability to perceive sound patterns within the environment, they begin to use this same energy channel to transmit information. "This generates interest in producing the sounds for detection by conspecific organisms" (Gans 1992, p. 9). This process is modulated by the environment, which propagates, reflects, absorbs and modifies sounds in different ways. In evolutionary terms, sound perception and production broaden the possibilities of interaction between individuals, as well as between individuals and the environment. But this process does not happen without costs. For example, a successful communication between individuals of the same species depends on tuning the sound patterns exchanged between sender and receiver. The energy expended in this process represents costs of various kinds, including "a risk of informing predators of one's existence and location" (idem, p. 10).

It seems reasonable to imagine that sounds vocalized by an individual should be adequately perceived by a fellow member of their species or by themselves, although this cannot be generalized. Exceptions, even when they serve to confirm the rules, are always instructive. It is known, for example, that the auditory system of toads and frogs has developed in a peculiar way and that many anurans do not even have a tympanic membrane like ours. Their hearing is produced by two organs, the amphibian papilla, sensitive to low and mid-frequencies (typically 50 Hz to 1 kHz), and the basilar papilla, sensitive to higher frequencies (above 1 kHz) (Goutte et al. 2017, p. 1). Even in anurans where tympanic middle ears are not present, sounds are sent to the inner ear through cavities or bones, making the animals able to hear the frequencies of their own calls.

But in some anuran species such as Brachycephalus ephippium and Brachycephalus pitanga, the poorly developed auditory system is apparently unable to capture the frequencies emitted by individuals of the species. This behavior poses a challenge to the conception that, in the evolutionary process, actions without a purpose represent a waste of energy and tend to be dampened. In a recent study, this behavior was interpreted as an aspect of a particular moment of evolutionary transition, in which the 'silent' calls have not yet been overcome, despite the ineptitude of the ears of that species. Since "signal production is energetically costly and sound may attract predators and parasites" (Goutte et al. 2017, p. 4), these high-pitched calls would have been maintained as a side effect of the visual aspect that accompanies the production of these sounds due to an "evolutionary inertia" (ibid).

One might speculate that not all of our actions (and not all of the frogs' actions as well) can be fully explained in terms of evolutionary features, such as scaring off predators or attracting sexual partners. After all, it is not because we ever murmured an imaginary song during a lonely walk that we would be going against the evolutionary process of the human species. Feeling the vibration of our vocal folds, creating melodic waves that resonate in our heads and thorax may be regarded as a pleasing and powerful way of connecting ourselves with our own bodies and with the world around us. This is by no means a waste of energy.

2 Sound as Image

In terms of acoustics, sound is an oscillatory movement of molecules in a material medium. Different vibratory patterns would lead us to perceive different sounds. As a vibrational phenomenon, the occurrence of sound is independent of, although it is closely related to, our perception of it. When we hear Beethoven's music or someone's voice, our brain produces a sound image that is triggered by the oscillating motion of molecules surrounding us. But if we refer exclusively to the acoustic phenomenon, we may find it difficult to identify what we understand as Beethovenian or vocal aspects in these oscillations.

It can be argued that the word "sound" can refer to quite different meanings, which often ends up blurring the distinctions among fundamentally different aspects of sound production and perception. Take, for example, the classic question: If a tree fell in the forest, but no one heard, would there be a sound? This philosophical puzzle makes sense because we have a single term to represent two different phenomena: the acoustic vibrations caused by the falling tree and the perception of these vibrations. In fact, what we understand as sound is the result of different levels of interaction, among which mechanical vibration is just one. Indeed, if we were to use two distinct terms to indicate the acoustic movement of material particles and the auditory images they produce in our minds, we could dispel some misconceptions about sound. As we shall see in this text, we intend to consider sound not as the object of listening, but as its medium.

The relationship between the acoustic production of sounds and the aural images they generate is generally taken as a direct cause and effect relationship. In fact, the perception of sounds is modulated by several factors ranging from the sociocultural contingencies to the cortical processing of the received signals to the listener's emotional state. Context plays an important role in the way we understand sounds. On the other hand, sounds can considerably influence other senses and even direct

our actions. Just think about the way sounds are produced in cinema to understand how the interaction between vision, hearing and semantic contexts can guide our hearing: For example, in a film's sound design, recording the tapping on a block of wood will make the gallop of a horse more realistic than the sound recording of an actual horse.

To give another example, an adequate DJ playlist can successfully encourage more individuals to dance at a party, just as the appropriate choice of musical repertoire can help with the consumption of dishes in a restaurant. Recent research shows that a soundscape can enhance certain taste characteristics of food, making customers willing to pay more if the meals come with a satisfying sound experience (Carvalho et al. 2015). Other experiments have shown an association between sour-tasting foods with high pitches and bitter flavors with low pitches (Crisinel & Spence 2009).

Diana Deutsch, an eminent researcher of musical illusions, carried out an experiment that challenges the notion that voice and listening operate separately. Deutsch realized that much of what we hear is not only related to the acoustic signals that reach our ears, but to the context in which the listening takes place and to our individual aural experience. Although she was interested in listening illusions, and therefore in unusual situations of sound perception, some of Deutsch's experiments on speech comprehension show that in voice listening, "the words and phrases that we hear are strongly influenced not only by the sounds that reach us, but also by our knowledge, beliefs, and expectations" (Deutsch 2019, p. 104). In demonstrating what she calls phantom words, Deutsch uses recordings of repeating words in a loop, over and over again, through stereo loudspeakers. Some of these words are just two-syllable words repeated out of phase on both channels. After listening for a while to those repetitions, listeners tend to hear different words, "nonsense words, and musical, often rhythmic sounds, such as percussive sound or tones" (idem, p. 106)[3]. The most impressive aspect of her experiments is that individuals perceive words in different languages and even with different accents depending on their original language and culture.

In this text, we seek to highlight the relational character of voice in different dimensions. First, in its connection with listening, establishing a relationship between receiving sounds and producing sounds. Second, pointing out that this relationship is always mediated by a body. This is not an abstract, idealized body, nor a body reduced to its carnal, physiological condition: The body here represents the empirical existence of a subject – and, eventually, of a thing – in time and space. Third, this existence implies the articulation between subject and the world, not as two separate characters, but as two perspectives of the same process. Thus,

Fig. 1: *Chorus II* (1988), framed black and white photographs, 142 x 188 cm, by Christian Marclay.

the body, when vocalizing, puts its surroundings in resonance, producing echoes that return to the body itself. Both what is vocalized and what is heard belong to the same process. Likewise, body and environment operate as an interconnected system, sounding and resonating with each other.

3 Silent Voices

There are over 20 small frames containing black and white photos of open mouths silently singing or screaming. The pictures themselves are arranged on the gallery wall to form the outline of a large mouth. The photos, cut from different sources, allude to a great choir in which men and women of different races and times create a silent harmony. As in other works by artist Christian Marclay, *Chorus II* (1988; see Fig. 1) explores the expressive potential of sound, not from its acoustic components,

but from references that are built from sound images. *Chorus II* allows us to listen with our eyes and to understand that our senses do not work autonomously and isolated, but develop as a network in which stimuli, memories, experiences, actions and reactions are interconnected. The work also indicates the material dimension of our sensations. Seeing or hearing do not refer to abstract categories of interaction with the world, but are ways in which our bodies act on the world at the same time that they are constituted by what, directly or indirectly, acts on us. Despite the absence of physical vibrations, *Chorus II* does sound: Spectators build their listening from their experiences and memories of other voices, other mouths, other choirs.

Marclay ingeniously connects voice, body, and listening. The wide-open mouths in an almost aggressive attitude remind us of an identifiable, strong, powerful sound. We may internally hear a scream, a groan, or some sort of tense utterance. The image of the mouths is the image of the sounds of those mouths. As spectators we connect these two image categories based on our experience in listening and voicing. That is, every time a scream is heard, we immediately associate it with the energy that escapes from the body, with the muscle contractions in the vocal tract, with the sensation of the cheeks stretching to the limit of the skin. In short, the static image that represents a gathering of mouths invokes, albeit in an imaginary way, the world of vibrations.

The idea that these mouths have no bodies is perceived as a contradiction. We have learned through experience that vibration depends on the existence of a vibrating body that defines and is defined by a space. Or as Douglas Kahn reflects: “Vibrations through their veritable movement generated a structured space and situated bodies and objects in that space. This process of situating did not outwardly transform the bodies or objects themselves, however, it just placed them in an ever-dependent relation within a larger system” (Kahn 1992, p. 15). Thus, the voices in Marclay’s *Chorus II* are not disembodied ones. On the contrary, they forge their bodies as extensions of their mouths.

In some of Marclay’s silent objects the connection between speech and listening becomes evident. He has created different works exploring the conditions posed by telephone devices in which transmitter and receiver are mounted in the same block, making the regions of voice production and listening geographically close. In these works emerges what Kahn calls residual sounds. These are not sounds that we expect to hear from familiar sounding objects, but sounds that “remain closed secured” in the stillness of the objects (Kahn 1994, p. 23). Despite the limitations imposed by its “physical, phenomenal silence, [...] a residual sound may be incredibly raucous” (ibid).

Chorus II resonates the modern imagery of the voice mediated by all kinds of sound devices: From the first phonographs and radio, to cinema, to current digital music streaming, we are used to hearing disembodied voices and we have learned to recompose these bodies in our imagination. Samuel Becket's *Not I* (1972) is exemplary in this regard. Minimalist in form, the piece, in which only the actress's mouth is illuminated by a light spot, is an explosion of auditory images. It is from the voice that the audience composes the character. The absence of the body channels attention to the voice, and it is this attentive listening that triggers our experiences and expectations to build an image for the hidden character. On the other hand, the mouth, which remains illuminated, insists on reminding us of the presence of a vocalizing body.

To some extent, *Not I* predates the current listening condition in which sound sources are usually hidden behind the membranes of speakers, headphones, and other sound devices. French composer Pierre Schaeffer named this condition acousmatic listening. Schaeffer was especially interested in discussing the role of listening in relation to his proposal for a musique concrète, a composition produced from sounds collected, recorded and arranged on a physical medium such as a disk or magnetic tape. For Schaeffer, the acousmatic situation favored what he called reduced listening: By hiding the sound source, the listener could focus their attention on the sound qualities, disregarding all the references sound could provide (Schaeffer 1966). Obviously, it is impossible to forget that behind the speaker – the acousmatic curtain that hides the sound sources – there is, or once was, a source. It is the sources that the loudspeaker conveys; whether it is a vibrating mechanical body, the electronic circuitry of a synthesizer, or the abstractly generated bits in a musical software, we cannot get rid of what is concealed by the speaker. Our conviction that there is something on the other side of the loudspeaker prompts us to reconnect the sounding objects with their vibratory movements and the way they sound. We infer that every object has a sound, a voice, just as every sound, every voice, comes from an object, from a body. This association between sounds and the objects that produce them, however, is the result of our experience, our expectations, our history and the history of the objects that sound. During our life we repeatedly listen to the voices of things and we build, in our memory, an association between things and sounds. These associations can assume different natures. We know that the sounds produced by large bodies tend to be lower than those produced by small bodies. We also learn to repudiate certain sounds (the noise of the upstairs neighbor) and to fear others (the sudden rumble of thunder during a storm). In a similar way, we go into alert when we hear a fire engine siren or the beep indicating the arrival of a message on the cell phone. Despite the strong connection

between sound as a vibrational phenomenon and the aural images it may provoke, this connection is neither stable nor unique, but relational.

Nina Sun Eidsheim, an academic dedicated to critically exploring the possibilities of voice representation, refers to the acousmatic situation to highlight this relational and multidimensional character of voice. It is assumed that when listening to a voice, we can learn something about the speaker, even when they are not visible to us. In this acousmatic situation, we are urged to ask a fundamental question: Who is this? Who is speaking? This fundamental 'acousmatic question' is based on the premise that there is a direct and stable relationship between sound and its source and that if we pay attention to the sound of a voice, we will be able to recognize the speaker, learn about their personality and even about their mood. However, Nina Eidsheim argues that there is no stable answer to the acousmatic question and that it arises precisely because of the "impossibility that the question will yield a firm answer" (Eidsheim 2019, p. 3).

The impossibility of answering this acousmatic question comes from the fact that neither the voice nor the vocal tract are static systems: They are subject to physical, emotional and contextual conditions that involve both the vocalizer and the listener. They both operate dynamically in the construction of meanings that emerge from vocal production. The inferences a listener may produce regarding a speaker's physical, racial or gender characteristics based on their voice is part of a process in which vocalizer and listener are both involved. For example, gender may be signaled by vocalizers "through word choice, intonation, speed, rhythm, prosody, level of nuance" (idem, p. 6-7), while listeners will "bring gender expectations to the vocal scene" (ibid). Eidsheim summarizes this complex interaction between actors of voice production/reception and its context in three corrective statements: "Voice is not singular; it is collective. Voice is not innate; it is cultural. Voice's source is not the singer; it is the listener" (idem, p. 9).

Eidsheim takes her own experience to discuss the relational status of voice. Despite her Korean origins, she was raised in a small town in Norway where her Korean identity was never a salient issue. On the other hand, when she visited Seoul, she realized that people treated her as a foreigner, despite her Asian traits. Eidsheim recalls that during her singing training in Norway she "participated in master classes offered by well-known American voice teachers" (Eidsheim 2008, p. 28) and that they "had been puzzled by [that] Asian-looking girl who spoke Norwegian and who, to their surprise, possessed a signature Nordic classical timbre" (ibid). A few months later, this time in California, a teacher complimented her on the quality of her voice, adding that her timbre was "really quite characteristically Korean" (ibid).

Eidsheim's experience poses a question: If the timbre of a voice represents an identity, a signature, how could these different situations produce such different perceptions of her cultural and ethnic identity? Starting from her own experience, Eidsheim develops the argument that, contrary to the current idea that the voice is "an unmediated manifestation of the body" (idem, p. 30), it "is indeed mediated" (ibid). Therefore, voice perception is conditioned by these elements of mediation, which are not only physiological, but cultural, social, subjective and contextual. Eidsheim is particularly interested in deconstructing the idea that it would be possible to hear race from the timbre of a particular voice. In this case, vocal training – that is, the way we use our bodies and voices – would be more significant for the perception of vocal timbre than any physiological differences linked to specific ethnicities. Thus, although an uttered voice presents indices of the uttering body – as an individual (gender, social status, age, etc.) and as a spatial position (Bertau 2008, p. 101) –, these signs are constructed in an intersubjective way, among the individuals and based on their interactions. The reliability of these indices in revealing speaker characteristics has attracted the attention of many researchers (Pisanski & Bryant 2019). Jody Kreiman and Diana Sidits make a significant contribution to understanding how the voice may (or may not) offer clues to the recognition of characteristics such as physical size, sex, age, health, appearance, racial group or ethnic origin of a speaker (Kreiman & Sidtis 2011). The authors emphasize that it is necessary to distinguish 'learned' from 'organic' marks, as they separate what the speaker can modify from what is subject to their physiological and anatomical constitution (idem, p. 111). It is also important to distinguish between marks and stereotypes of speaker characteristics. While marks are generally "reliable cues to that characteristic", stereotypes are related to what "listeners expect to hear from a speaker who possesses certain physical attributes" and, therefore, "social expectation influences listener's judgments". At the same time, these stereotypes contribute to "vocal behaviors children learn as they grow" (ibid). Thus, associating voice with an individual's specific characteristic is not a trivial issue. For example, while the distinction between male and female voices can be achieved with some consistency, to transgender individuals, there is an important interplay between organic and learned characteristics. Since voice is an important index of an individual's social and personal characteristics, "producing a female (or male) sound with what remains a male (or female) vocal tract and larynx" (idem, p. 144) can be a challenging demand for individuals who have undergone transgendering surgery. Thus, what the voice can say about an individual depends on the balance between what is physically and physiologically determined in the production of voice and the intersubjective experiences that make up our listening

Fig. 2: Hologram of Hatsune Miku at a live concert.

processes. When we listen, we project our knowledge, our beliefs and expectations onto the speaker. Therefore, we always hear a little of ourselves in the 'other'.

4 Disembodied Voices

As is the case with the beginning of any pop concert, the band attacks the first chords, the lights come on, while fans wait anxiously for the entrance of the main character, the singer. In contrast to the other musicians in the band, in this case the singer is not a person, but an avatar projected holographically onto the stage. Her blue hair and high school student clothes make explicit reference to Japanese manga. Hatsune Miku reproduces the cliché image of what a popstar should be. Initially restricted to otaku[4] circles, she gradually attracted the attention of other musical circuits. Without assuming polemic attitudes or wearing extravagant clothes, Hatsune Miku[5] became an iconic representation of Japanese pop culture. Thanks to Vocaloid[6], a software that produces singing voices artificially, Hatsune Miku has become a virtual idol.

From extensive sound banks, Vocaloid allows its user to type in the lyrics of a song and synthesize it from a series of instructions. In essence, the Vocaloid interface superimposes the song text over a kind of musical score. A series of subtle adjustments can be applied to each vocal sound, allowing the creation of very sophisticated vocal articulations. Miku is probably the best known of a series of virtual artists produced with the help of Vocaloid. In part, her success is due to the fact that graphic projections added a visual image to the singer's well-behaved voice. Other software such as MikuMikuDance[7] made it possible for fans to import 3D models and create their own animations of the singer. In a short time, what was

seen was “a boom in user content and the development of other imitated characters” allowing that “fans’ animations become part of the concerts” (Bessant 2018, p. 31). Encouraging the use of these programs by amateur musicians and animators ended up creating a sense of community that was built around very specific aesthetic and cultural values.

Hatsune Miku draws attention because of a contradiction that is inherent in her existence. On the one hand, she is recognized as representing a certain category of singers with whom she shares certain similarities – dressing habits, musical genre and, most importantly, a (professional and expressive) singing voice. On the other hand, like Frankenstein’s creature, Hatsune Miku is relegated to being an outsider, a mirror of all female singers without being any of them. She does not go to parties, does not have a boyfriend, does not donate to social and ecological causes. Hatsune Miku does not issue opinions. Her voice is doomed to be essentially what she was designed to be: a general voice, unbiased and flawless.

Miku’s synthesized voice poses a problem that is part of the growing mediation process to which our bodies and our senses are submitted. Speech and listening interfaces, like other devices that surround us, are perceived as interfaces produced to enable our interaction with other individuals or with other devices in a neutral way. What would be understood as intentionality or as attributes of the agents taking part in a process of social interaction – a conversation, a love relationship, a dialogue between teacher and student – is often perceived as an accidental contingency when a technology is at issue. Thus, the compression of mp3 files, the noise produced by the hair dryer or the limit of bandwidth in a phone conversation, are not seen as choices, intentional or not, of those who produced these devices, but as something that is part of their ‘nature’. Ideally, we can imagine that the sounds we get from our headphones and the voices Siri employs to communicate with Apple users are designed to sound generic and to be adapted to any situation. Sound technologies appear as if it were possible to invent a generic voice aimed at generic listening. Thus, a synthesized voice is founded on the belief that it is essentially neutral and therefore can be shaped to take on any character we wish. However, this neutrality clashes with what we perceive in our daily experience, in which both voice and listening are subject to physical, emotional, and cultural conditions, making a voice always something unique, referring to a field of experiences which are modulated by a specific act of listening.

In 1955, Max Mathews joined Bell Laboratories as an acoustic engineer to investigate efficient ways of transmitting and receiving voice over the telephone. In the following years, Mathews’ research would unfold into a series of breakthroughs, not in

communication but in music technology. His achievements lead to what is currently known as computer music. In 1957, he created a computer language called MUSIC to produce sounds, and from then on he was the protagonist in a series of inventions capable of generating or controlling sounds electronically. If the first digitally produced sounds in 1957 did not excite Bell's engineers, a few years later, in 1961, Mathews and his colleagues were already able to reproduce comprehensible vocal sounds. *Daisy Bell* (*A Bicycle Built for Two*), an old folk song reproduced entirely by sound synthesis, was impressive enough to be used by Stanley Kubrick towards the end of his 1968 film *2001: A Space Odyssey*. In the final clash between man and machine, the HAL 9000 computer 'sings' *Daisy Bell* as his memory is disabled. Like the voices offered by Vocaloid, the synthesized version of *Daisy Bell*[8] produced at Bell Labs does not come from a real body. However, they lead us to construct coherent images of possible bodies (a teenage Japanese popstar like Hatsune Miku or a fictional powerful computer like HAL 9000). Unlike Marclay's *Chorus II*, in which we are invited to create voices from images of vocalizers, in these processes of vocal synthesis we take the opposite path to imagine who these voices could belong to. Both the voice as sound and the vocalizer as an individual emerge as images provided by our experience as listeners. In fact, no voice, be it natural or artificial, is neutral, not just because it supposedly belongs to a subject, but because we project our listening experience onto the voices we hear. And just like the vocalizer voices what they can hear, artificially projected voices are calculated from someone's idea of a voice. As we interact more and more with disembodied voices heard through loudspeakers, we are getting used to the idea that these voices represent a general idea of voice: accentless voices that could belong to anyone and no one at the same time.

However, HAL 9000, similar to today's virtual voice assistants like Apple's Siri or Amazon's Alexa, represents a series of values that are specific to a culture. Its generality is built from an imaginary idea of what is common. Disguised in the form of algorithms or electronic components, this generality actually results from the subjective view of those who implemented it. By creating a general self, these technologies rule out the existence of an 'other': Virtual voice assistants are not designed to understand minority discourses, immigrant accents, or social groups that express themselves in slang and dialects. That is, they understand those who speak within what is set as standard. As the access to a significant number of services becomes dependent on voice recognition, the supposed neutrality of recognition algorithms becomes a political form of segregation that has only recently been highlighted by authors such as Safiya Noble (2018) and Cathy O'Neil (2016). Current electronic speech production and electronic speech recognition, like any

other computational procedure, depends on the implementation of models in the form of algorithms. Models always represent a simplified view of a problem. Models have a purpose, seek to represent certain aspects of reality and are often evaluated in terms of efficiency, generality or accuracy. However, models are based on choices and always represent a point of view. When implemented in the form of an algorithm, they reproduce the expectations, habits, beliefs and, eventually, the prejudices of those who created them. These voices are therefore expected to act as references of dominant cultural and social strata and help construct references of what would be a good vocal quality and appropriate modes of vocalization.

5 Incarnated Voice

Luiz Ernesto Machado Kawall is a Brazilian journalist and museologist passionate about voices. Born in 1927, he has devoted much of his life to collecting voice recordings. His collection, created with his own resources and driven by his own curiosity, brings together some ten thousand voices recorded in different contexts: from famous people, to ordinary individuals, to animal voices. In addition to its historical value, this collection enables an experience that we could hardly reproduce in our daily life: listening, one after another, to voices from different times and places, imagining the situations in which they were recorded and the bodies that produced them. His vozoteca[9] (voice library) is an opportunity to perceive through listening the diversity of speeches, accents, languages, subjects, and musics that came into being through voice. This diversity contrasts with the restricted set of voices that we access in our daily lives. Kawall's collection gives the dimension of the limits of our own listening. In our day-to-day social life, our contact with people who speak other languages, who express themselves with other vocabularies, who belong to different social groups, is quite restricted. Kawall sought to overcome these limitations by seeking access to voices that were impossible to hear and record. He became interested, for example, in hearing the voices of characters who were never recorded, such as Dom Pedro I, emperor of Brazil in the early 19th century. Kawall was also a collector of cordel, a form of popular literature, done informally and printed in pamphlets. Perhaps it was the proximity to the cordel that aroused his interest in a recurring character in this form of literature: Virgulino Ferreira da Silva, better known as Lampião.

Leader of a gang that operated in northeastern Brazil, Lampião was considered the king of the "cangaço", a term for a movement of bandits who acted against government and paramilitary forces in very arid and poor regions of the country. Transformed into a popular hero who fought against police oppression and

wealthy farmers, Lampião became an iconic character in the rural history of Brazil. Since he lived nomadically and clandestinely in remote regions of the country and was killed in 1938 by police officers, it is unlikely that his voice was ever recorded. In his obstinacy to know the cangaceiro's voice, Kawall ended up resorting to Umbanda, a syncretic religious tradition that brings together elements of Catholicism, the tradition of African orixás and spirits of indigenous origin. Some Umbanda practitioners called "aparelhos" (literally "devices") are mediums who have the ability to embody spirits, allowing them to communicate with human beings. Kawall visited one Umbanda temple in which a medium incorporated the spirit of Lampião, enabling him to record the voice of the deceased cangaceiro[10].

Fig. 3: Lampião, 1927 (photographed by Benjamin Abrahão Botto).

One might ask: Whose voice was registered by Kawall? To what extent are those recordings representative of Lampião? These questions bring us back to the impossibility of providing a definitive answer to the acousmatic question posed earlier by Eidsheim. In this case, there is no simple answer. The acoustic voice recorded by Kawall was not provided by Lampião's body, but by a transducer – in this case, not a loudspeaker, but a medium who incorporated Lampião's spirit. Listening to Lampião's voice involves beliefs and subjectivities as these recordings only make sense in the set of representations and experiences of those who listen to them. Lampião himself cannot be reduced to an individual's physical body, nor his voice to an acoustic trace. His presence was built in the symbolic imagery of rural culture in Brazil from a network of reports, beliefs, and fantasies established through speech and listening, orality and aurality. In fact, it may be irrelevant to know how much of this knowledge corresponds to what he actually was as an individual. By adding another element to this complex imagery that represents Lampião, the voices recorded by Kawall are as true, as real, as the stories told in prose and verse by the popular culture of cordel in northeastern Brazil.

Without the presence of his body, and without the existence of an acoustic signal coming from that body, the voice recordings made by Kawall may or may not be taken as realistic. It completely depends on our intention and ability to listen to them.

6 Resonances

> "Sound is not what we hear, any more than light is what we see" (Ingold 2007, p. 11). The coherence of the association between sound and image is rarely questioned. The profound asymmetry between these two entities is based on the fact that the objects we see come to us through light, and the objects we hear come to us through sound. Just as we do not see light, Tim Ingold warns us that sound "is not the object but the medium of our perception. It is what we hear in" (ibid). Sound is what makes listening possible, but it is not what we listen to. In dialogue with Ingold, composer and scholar Rodolfo Caesar continues: "sound is the support/transport [...] that allows us to listen to 'sound images': sound objects, sonorities, words, etc. Just as light provides us with the exercise of visuality" (Caesar 2020, p. 85).

Thus, voice is also a medium rather than an object. Like its counterpart, hearing, voice is not a thing, but a relation established between subjects and objects, between what is inside and what is outside. This relationship can be understood as resonance, the ability to sound from – or with – the energy produced by an 'other'. When I speak, I resonate in someone else's listening and in my own listening. I am both a sounding subject and sounding object. Or, as Steven Feld puts it, "one hears oneself in the act of voicing, and one resonates the physicality of voicing in acts of hearing. Listening and voicing are in deep reciprocity, an embodied dialog of inner and outer sounding and resounding built from the historization of experience" (Feld 2003, p. 226).

When we are born, we begin to establish this experience. At this starting point, voice and listening, subject and object, are one and the same. Our voice/listening is formed as an imitation, a mirror of what we hear from our mother. At some point, the lullaby we hear becomes the cry we emit (Clough 2013, p. 66) and the annoyance caused by our crying establishes the channel of communication with an 'other'. What is built is a resonance, the possibility of sounding and hearing in a feedback process, the "delicate looping that is listening or being heard" (ibid). From then on, our vocal folds and our eardrums become inseparable membranes.

References

Bertau, Marie-Cécile (2008): Voice: A Pathway to Consciousness As 'Social Contact to Oneself'. In: Integrative Psychological & Behavioral Science. 42. Pp. 92-113.

Bessant, Judith (2018): The Great Transformation: History for a Techno-Human Future. London, New York, NY: Routledge.

Caesar, Rodolfo (2020): Apontamentos para espetar o som. In: MusiMid. 1/1. Pp. 82-87.

Clough, Patricia T. (2013): My Mother's Scream. In: Sound, Music, Affect: Theorizing Sonic Experience. Marie Thompson; Ian Biddle (eds.). New York: Bloomsbury.

Crisinel, Anne-Sylvie; Charles Spence (2009): Implicit Association between Basic Tastes and Pitch. In: Neuroscience Letters. 464/1. Pp. 39-42.

Deutsch, Diana (2019): Musical Illusions and Phantom Words: How Music and Speech Unlock Mysteries of the Brain. New York, NY: Oxford University Press.

Eidsheim, Nina S. (2008): Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance. PhD thesis. University of California, San Diego.

Eidsheim, Nina S. (2019): The Race of Sound: Listening, Timbre, and Vocality in African American Music. Illustrated edition. Durham: Duke University Press Books.

Feld, Steven (2003): A Rainforest Acoustemology. In: The Auditory Culture Reader. Michael Bull; Les Back (eds.). Oxford: Berg. Pp. 223-239.

Fogg, Thomas (2018): Expériences Sonores. Music in Postwar Paris and the Changing Sense of Sound. PhD thesis. Columbia University.

Gans, Carl (1992): An Overview of the Evolutionary Biology of Hearing. In: The Evolutionary Biology of Hearing. Douglas B. Webster; Arthur N. Popper; Richard R. Fay (eds.). New York, NY: Springer. Pp. 3-13.

Goutte, Sandra; Mason, Matthew J.; Christensen-Dalsgaard, Jakob; Montealegre-Z, Fernando; Chivers, Benedict D.; Sarria-S, Fabio A.; Antoniazzi, Marta M.; Jared, Carlos; Sato, Luciana A.; Toledo, Luís F. (2017): Evidence of Auditory Insensitivity to Vocalization Frequencies in Two Frogs. In: Scientific Reports. 7/1. Article No. 12121.

Ihde, Don (2012): Listening and Voice: Phenomenologies of Sound. Second Edition. Albany, NY: SUNY Press.

Iazzetta, Fernando (2016): A imagem que se ouve. In: Diálogos Transdisciplinares: Arte e Pesquisa. Gilbertto Prado; Monica Tavares; Priscila Arantes (eds.). São Paulo: ECA/USP. Pp. 376-395.

Ingold, Tim (2007): Against Soundscape. In: Autumn Leaves: Sound and the Environment in Artistic Practice. Carlyle Angus (ed.). Paris: Double Entendre. Pp. 10-13.

Kahn, Douglas (1994): Christian Marclay's Lucretian Acoustics. In: Christian Marclay. Daadgalerie: Berlin. Pp. 23-34.

Kahn, Douglas (1992): Histories of Sounds Once Removed. In: Wireless Imagination: Sound, Radio, and the Avant-Garde. Douglas Kahn; Gregory Whitehead (eds.). Cambridge, Mass.: MIT Press. Pp. 1-29.

Kreiman, Jody; Sidtis, Diana (2011): Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception. Malden, MA: Wiley-Blackwell.

Noble, Safiya U. (2018): Algorithms of Oppression: How Search Engines Reinforce Racism. Illustrated edition. New York: NYU Press.

O'Neil, Cathy (2016): Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown.

Pisanski, Katarzyna; Bryant, Gregory A. (2019): The Evolution of Voice Perception. In: The Oxford Handbook of Voice Studies. Nina Eidsheim; Katherine Meizel (eds.). New York: Oxford University Press.

Reinoso Carvalho, Felipe; Van Ee, Raymond; Rychtarikova, Monika; Touhafi, Abdellah; Steenhaut, Kris; Persoone, Dominique; Spence, Charles (2015): Using Sound-Taste Correspondences to Enhance the Subjective Value of Tasting Experiences. In: Frontiers in Psychology. 6. Article No. 1309.

Schaeffer, Pierre (1966): Traité des Objets Musicaux. Paris: Éditions du Seuil.

Tomatis, Alfred A. (1992): Conscious Ear. Barrytown, N.Y: Station Hill Press.

Notes

1 The idea of sound as image has been developed in Iazzetta 2016.

2 Tomatis' trajectory is quite controversial and includes the discrediting of his colleagues in France and the commercial use of his scientific findings, particularly his 1953 invention of the Electronic Ear, and later of his TOMATIS® Method. For a critical discussion on the French physician's interest in the sense of listening, see Fogg 2018.

3 Some phantom words examples can be found at: http://dianadeutsch.net/book_audio/Modules-2019/mixdowns/MP3/ch07ex01_phantom_words-d5_mixdown.mp3 [last accessed August 25, 2021].

4 A Japanese subculture interested in manga and anime.

[5] Hatsune Miku was released by Crypton Future Media in August 31, 2007.

[6] Released in 2004, Vocaloid allows its users to type text and melody to synthesize a song. Voice synthesis is performed using voice banks extracted from samples produced by professional singers. See https://www.vocaloid.com/en/ [last accessed August 25, 2021].

[7] MikuMikuDance, or MMD, is a freeware animation software originally produced to give life to the Vocaloid character Hatsune Miku. Since its launch in 2008 the program has attracted the attention of a wide community on the Internet interested in creating characters based on anime culture. Many MMD videos can be found on NicoNico, a Japanese video-sharing service on the web.

[8] The similarities between Vocaloid and voice synthesis produced at Bell Labs had already been noticed by Kenmochi Hideki, leader of the research project that gave rise to Vocaloid. In its first public appearance at Musikmesse in 2003, the program was to be called *Daisy*, in allusion to *Daisy Bell*. The name had to be changed for copyright reasons. See: Red Bull Music Academy 2014, https://daily.redbullmusicacademy.com/2014/11/vocaloid-feature [last accessed August 25, 2021].

[9] His voice archive was donated to the Institute of Brazilian Studies of the University of São Paulo in 2013 and can be accessed under the title Vozoteca LEK at http://200.144.255.59/catalogo_eletronico/ [last accessed August 25, 2021].

[10] These recordings are stored at the Institute of Brazilian Studies collections archives under the reference Vozoteca LEK, VOZ-CDr-010 at http://200.144.255.59/catalogo_eletronico/ [last accessed August 25, 2021].