Voice technologies are described experientially here, as the writer interacts with the multiple disembodied voices surrounding her and reflects on the often ambivalent and complex emotions they call up.

Finally, Martin Thomas's resonant piece describes how sound-capturing technologies have allowed Australian indigenous peoples to keep their knowledge and traditions alive, despite language loss, and to "write" their stories without having to inscribe them into a semiotic modality that divests them of their fundamental oral qualities.

1 Vox Humana: The Instrumental Representation of the Human Voice

Theo van Leeuwen

Introduction

I am a jazz pianist and church organ player as well as an academic. Two years ago I bought a new digital piano, a Roland RD300-SX. Its hundreds of voices imitate the whole range of traditional musical instruments, but also include "human voices": Aah Choirs and Oooh Choirs, Jazz Scat, Space Voices, and more. As a church organ player, I knew this was nothing new. Church organs have included *vox humana* ("human voice") stops for centuries. I became intrigued. Is there a continuity between analog and digital musical representations of the human voice, or has digitality introduced a new dimension? It is this question that I will explore in this chapter.

Writing about instrumental representations of the human voice, it is impossible not to touch on the question of authenticity. To do so, I will use "modality theory," an approach that asks not whether voices actually are "authentic," but *as how* "authentic" they have been represented. As it turns out, musical representations of the human voice have always been relatively abstract, perhaps in the first place seeking to provide a kind of discourse about the human voice, rather than seeking to be heard as realistic representations of human voices.

The Vox Humana

Already in the course of the sixteenth and seventeenth centuries, church organs were beginning to include voices representing other musical instruments, natural sounds, and also the human voice.

"Does it not suffice," wrote Pierre Trichet in his *Treatise on Musical Instruments*, published in 1640, "to have found a way that in the same organ

. . . one can hear the song of the lark, the twittering of the nightingale, the drone of the pedals ("bourdonnement"), the noise of drums, trumpet fanfares, the echoing of clerons, the clanking of cymbals, the tolling of bells, the buzzing ("nazardement") and grating sound of the régales or voix humaines; in short the sound of almost all musical instruments, such as the playing of soft flutes, fifes, flageolets, arigots, oboes, bassoons, cornets, cromornes, musettes, violins? . . . All these all together, if desired, or by combining them and separating them alternately."[1] And Father Mersenne, a friend of Descartes, wrote in his book *Harmonie Universelle* (1636): "If the violin stop is added . . . it seems that there will be nothing more to desire in the organ, unless it be that the pipes should sound the vowels and the syllables. This it seems must not be hoped for because of the great difficulty encountered."[2]

This vision of an instrument on which one player could command a whole orchestra continues to inspire instrument makers to the present day. And at the same time, another vision and another project emerged: the representation of the human voice by musical instruments. Organ builders began to create a new organ stop, the "vox humana" or "voix humaine," developing organ pipes that would imitate the human voice. For this they used reed pipes, which have within the pipe a resonating metal tongue that is wind-driven, and tunable by means of a clip that presses against it. Soon these pipes would be given a separate wind supply, so as to allow a kind of vibrato, by means of "tremulants," or valves covered in leather that were suspended in the wind conduit and started to flap as the wind entered the conduit, causing fluctuations in the wind supply, and hence in the sound of the pipes. The size and weight of the valves determined the rate and depth of the vibration.

Why did they use reed pipes, which sound rather thin and hollow, a little like an oboe, and not at all like a human voice? Perhaps the organ builders knew their classics. Reed instruments had been equated with the emotive aspect of music since antiquity. In the fifth century BC, the poet Pindarus told how Athena invented the *aulos* after hearing the lament of Medusa's sisters at the beheading of the Medusa. The aulos was a kind of oboe and would become the key instrument in Greek drama and tragedy because of its capacity for emotive expression through dynamic shading and tonal coloring. Reed instruments may not have sounded like human voices, but they were nevertheless conventionally associated with human

voices because of their emotive expressiveness. In the context of church organs, powered by large bellows and unable to vary the dynamics and timbres of their individual voices, this was a novelty.

But a close imitation of the human voice it was not, and the way the organ was played had to help make the imitation more convincing. Organists warned each other to play "compassionately and very legato"[3] and to make sure they did not use too wide a range. "The voix humaine must not be played lower than the first F ut fa, not higher than the fourth G re sol, because natural voices do not ordinarily pass these limits," wrote Dom Bedos de Celles in the eighteenth century, adding that using the tremulant was particularly important: "Without the tremulant the Voix-Humaine does not truly imitate the human voice. I know of only two Voix-Humaines which have achieved that quality. I believe the perfect success of these two Voix-Humaines must be attributed principally to the quality of the Tremblant-Doux."[4] And according to Mersenne, "such a 'Tremblant'" had to "beat in such a way that it imitates the vibrato of human voices in the stops of the organ," and for this reason it had to be "close to the [wind]chest, where the wind does not vibrate too fast."[5]

Music, Emotion, and the Voice

The Baroque period foregrounded the music of drama and tragedy, and, as Curt Sachs describes it in his insightful history of musical instruments:

> The new generation strove for strong emotion: it tried to appeal to the listeners' hearts. "What passion cannot Music raise and quell," sings Dryden. Father Mersenne relates that Italian singers of his time expressed passions so violently that the audience would think they themselves were involved; and when, in an opera by Monteverdi performed in 1608 at the Mantuan court, Arianna, forsaken by Theseus, sang her heart-rending lament, the listeners burst into tears.[6]

As a result, polyphony receded and homophony—a form of music in which a single expressive melody stands out against a background of chordal accompaniment—began its ascendance. Homophony was thought to be more suited to the expression of emotion than polyphony, especially in opera. This happened first of all in singing, but it also caused radical changes in instrumental music. Instruments that, like the human voice, could produce a wide range of pitch, dynamics, timbral variation, vibrato, and so on were kept and further developed, especially strings. Other instruments

disappeared in the space of a few decades—for instance, wind instruments in which the reed could not be grasped by the lips and that therefore did not allow "overblowing" in a higher octave, instruments such as the ranket, the cromorne, the schryari, and the rauschpfeife. The organ is also such an instrument, and organ builders sought to remedy its lack of expressiveness by creating voices like the "Viola da Gamba" and of course the "Voix-Humaine," putting them in a separate section of the organ, the *Rückwerk*, which was positioned behind the organist at the front of the organ gallery and looked like a mini organ all by itself. Voix-humaines also acquired their own windchests and their own, separate tremulants, and sometimes they were enclosed in thin-walled boxes to allow them to be played softer and louder, all to enhance the organ's potential for emotive expressiveness, and all for the divine instrument to become human—and for humans to become divine.

Developments of this kind have continued to the present day. The keyboard continues to reign and to be developed as a kind of master instrument—a compendium of all musical and other sounds. In the late nineteenth century, when Dutch clergyman Ten Kate heard in the organ of his church "the shepherd's flute, the wind, thunder, rain, streams, the nightingale, storms,"[7] the first theater organs had already been built, allowing an even wider range of sound effects, including doorbells, steamboat whistles, railroad bells, sirens, bird calls, and car horns, as well as many different musical instruments.[8] Wurlitzer theater organs also introduced the "double touch," whereby another voice could be introduced by pressing the keys harder, and the "pizzicato" touch, for a staccato or plucking effect. In 1927, one company, Marr and Colton, began to build organs with "symphonic registrators," presets with labels such as "Love (mother)," "Love (romantic)," "Love (passion)," "quietude," "jealousy," "hatred," "anger," "excitement," "agitation," "mysterious," "gruesome," "happiness," and "sorrow."[9] The violin and the cello continued to be key solo instruments and key instruments for the expression of emotion. In the twentieth century, the saxophone would play a similar role in jazz. Using elements of brass and woodwind, and combining the key arrangement of the oboe with the mouthpiece of the clarinet, the saxophone has an unrivalled capacity for the expressive variation of timbre, ranging, as Sachs put it, "from the softness of the flute, through the mellow tone of the cello, to the metallic strength of the clarinet."[10]

My own digital piano, the Roland RD300-SX, follows the tradition. Many of its voices imitate musical instruments, and others represent natural sounds, such as Seashore, Rain, Thunder, Wind, and Stream; animal sounds, such as Bird, Dog, and Horse Gallop; and mechanical sound effects, such as Cars, Trains, Planes, Gunshots, Machine Guns, and Explosions. Voice effects such as Laughter, Applause, and Screaming are also included. Like eighteenth-century European church organs, which featured different kinds of human voice stops such as the German *Jungfernregal* (Girls' Voices) and *Singendregal* (Singing Voices), and the Spanish *Viejos* and *Viejas* (Voices of Old Men and Women), *Vox pueri* (Boys' Voices), and *Vox tauri* (Bulls' Voices),[11] the Roland piano includes a range of human voices—Female Choirs, Aah Choirs, and Ooh Choirs, Humming, Jazz Scat, Solo Vox, Synvox, Analog Voice, and Space Voice. It also has a "humanizer" effect that can be added to every single one of the instrument's 270 voices, adding "a vowel colour to the sound" and "making it similar to the human voice."[12] Father Mersenne, living in a mechanical age, had not been able to foresee how the difficulty of producing vowel sounds and syllables might ever be overcome. Two-hundred and fifty years later, in the digital age, his vision was finally achieved. Or was it?

Abstract and Sensory Modality

To discuss how the modern Roland piano imitates the human voice, I will need to briefly recapitulate the theory of "modality" I introduced in an earlier publication.[13]

The idea of modality originated in philosophy as a term for the study of the truth conditions of propositions. When it was taken over in linguistics by linguists such as Lyons and Halliday,[14] an important shift occurred. While philosophers asked, "How likely is it that this proposition is true?" linguists asked "*As* how likely has this proposition been represented?" Linguistic signifiers of probability, such as the modal auxiliaries ("may," "will," "must") can be, and are, equally used for true and untrue propositions. Untrue propositions can be affirmed with great certainty, and, as evidenced by the case of the Holocaust deniers, doubt can be cast on true propositions. Language alone cannot guarantee truth, and linguistic analysis can therefore only bring out *as how* true something has been represented.

The idea of modality as apparent truth or verisimilitude has subsequently been applied to modes of communication other than language, including images[15] and can also be applied to representation by means of sound and music—for instance, the representation of musical sounds, natural sounds, and speech by musical instruments such as the Roland RD300-SX. In the case of sound and music, degrees of verisimilitude are signified by the degree to which certain means of auditory expression are used. Pitch range is one of these. You could, for instance, represent rain by means of a "pitter patter" rhythmic pattern on pizzicato strings with a very narrow and restricted range of just two pitches. This would not be very realistic, but it would nevertheless get the idea across. It would be a relatively abstract representation. You could also use an appropriate electronic sound with a good deal of hiss, increase the pitch range, and include dynamic variation and glissandos to suggest sheets of rain. This would not only be more naturalistic, but would also have a greater emotive effect. It would not only get the idea across, but also "feel" more like rain. And if the range of pitch, and of dynamic and timbral variation, is extended even further, the experience can become "more than real" and emotionally overwhelming. The cline of verisimilitude is not just a cline of verisimilitude, but also a cline of emotional involvement, a cline that runs from relatively "abstract" representation, through "naturalistic" representation, to what Kress and I have called *sensory representation*.[16] In the case of abstract representation, the truth criterion is conceptual: "to which degree does this sound represent the essential, constant qualities of what it represents, e.g., the 'idea of rain'?" In the case of naturalistic representation, the truth criterion is perceptual: "to which degree does this sound represent what we would have heard if we could have heard the real thing, under specific acoustic conditions?" In the case of sensory representation, the truth criterion is emotive: "to which degree does this representation engender the feelings that accompany an intense experience of what is represented?"

The same auditory parameters serve not only to indicate degrees and kinds of truth in representation, but also to indicate what Martinec[17] has called "presentation," for instance, the presentation of the identity of a speaker or a particular kind of action or event. In speech, a limited pitch range, for instance, is often used to present the "*not* human," such as the divine or the mechanical, as in stylized forms of ritual chanting or the monotonized speech of robots.

The key indicators of modality include not only pitch range, but also dynamic range. Instruments that do not allow dynamic variation are again a touch more abstract, a touch less emotive. As we saw, the organ is such an instrument, and eighteenth-century German organ builders therefore looked for ways to make the *vox humana* the only stop capable of dynamic variation.

Degrees of fluctuation form a third parameter, ranging from completely steady sounds to a maximally deep and/or rapid vibrato. Vibrato, as long as its speed and depth resemble the fluctuations the human voice is capable of, is a very effective signifier of emotive expression. Emotions can make our voice waver and tremble, and for this reason tremulants or tremblants played a key role in creating convincing imitations of the human voice in church organs.

Degrees of friction constitute a fourth parameter. "Friction" here refers to anything we hear besides and in addition to the tone itself—friction, hoarseness, harshness, rasp, sibilance, breathiness. This again increases verisimilitude. The guitar sounds of the Roland piano, for instance, lack the sound of the plectrum and of the fingers sliding along the fret board. This makes the sound inevitably less realistic and a little more abstract. Unlike musical instruments, the sound of the human voice is affected by different emotive states, and also by physical states such as fatigue, intoxication, health, and age. Different kinds of vocal friction can evoke different meanings, different emotions, and different associations. A rough, hoarse voice can signify having lived a rough, hard life. A breathy voice can signify exertion or excitement, and, combined with a soft whisper, it is a standard signifier of sensual seduction.

Degrees of absorption should also be mentioned. They range from the completely "dry" to the maximally spacious sound. Beyond a certain point, reverb can create an effect of dread, as in representations of extraterrestrial landscapes.

These parameters are not only capable of fine-grained degrees, but may also be combined in many different ways, providing a rich range of resources for the creation of verisimilitude and the expression of emotion. It is perfectly possible that some parameters of the same sound are "abstract" and others "sensory." The reed pipes of the church organ, for instance, are a fairly abstract and not very realistic representation of the human voice. At best, they put across the idea of a human voice, but they do not really

sound like one. The sound of the tremulant, on the other hand, is both more realistic and more able to pull the heartstrings.

Musical representation always combines the abstract and the sensory. In Honegger's *Pacific 231*, for instance, the sound of the locomotive engine is played by brass, strings, and timpani, alternating between two pitch levels, first low and slow, then gradually faster and higher as the machine gathers speed, and the whistle is played by strings abruptly sliding up. There is no friction, no grinding of wheels on rails, no squeaks, no hissing, no shrillness in the whistle. The music gives the idea of the locomotive and the whistle, but it does not aim at realistic representation. At the same time, it is music, and therefore emotively engaging. It does not only represent the machine, it also makes us *feel* its power as the energy level of the music increases. "What I was looking for," said Honegger in 1924, "was not so much to imitate the sounds of the locomotive, but to translate visual impression and physical enjoyment with a musical construction." [18]

The Human Voice According to Roland

The instrumental basis of the "human voices" in the Roland RD300-SX remains that of a musical instrument. In some voices, it in fact reminds of the thin, reedy sound of the old *vox humana*. As in the church organs, it is the added parameters, not the tone itself, which makes the voice "human."

In the Aah Choir, for instance, a vowel sound is added to the instrumental sound, but only as a kind of flavor. Although speech synthesis or sampling techniques would make it perfectly possible to use an actual "aah," this is not what Roland has provided. The voices of the Aah Choir do not aim to "be" human voices, only to represent them in a way that—like Honegger's representation of a locomotive—is both abstract, and, because the Roland remains a musical instrument, sensory. Also added is "friction," in the form of a hiss. Again, it is not clear what this represents. Breath? It is difficult to say. Perhaps it forms a relatively abstract representation of the idea of friction, the idea of the "grain of the voice."[19] Finally, there is a slight and slow vibrato, to suggest the fluctuating timbre of the human voice. As in eighteenth-century church organs, these effects work only in the middle two octaves. They must be complemented by the way the instrument is played.

The Ooh Choir too has a quite thin and reedy instrumental base, sounding, perhaps, somewhat like strings. The "ooh"-like vowel sound is stronger here and has a consonantal attack which, if played forcefully, staccato, and not too slowly, makes for a relatively convincing "doo" (sustaining the tones destroys the illusion in less than a second). Unlike the Aah Choir, hiss is absent here and the vibrato is slight. In other words, the Ooh Choir relies more on its vowel sound and less on friction and fluctuation. Again unlike the Aah Choir, this voice also works at two octaves below the middle C, producing relatively convincing male voices.

The Jazz Scat voice makes use of the piano's "humanizer" effect, which alternates different vowels (this can be done at twenty-one different frequency levels, making it hard to use). Different pressure produces different vowels. A soft touch yields "doo doo doo," more pressure produces "pah pah pah" and sometimes a fast descending glissando. Some of the variation is preset and hard to predict for the player. To create a convincing instrumental imitation of scat singing, players quickly learn to restrict themselves to a fairly circumscribed pitch range, tempo, and attack (again, only staccato). Only a stereotyped version of scat singing can be produced in this way, and, as with so many imitative voices, it is difficult to use the voice for anything other than clichés.

The Space Voice, finally, uses the same "ooh" sound as the Ooh Choir and also has a hiss. But its fluctuation is fast and shallow (therefore "*not*-human") and its reverb strong. This voice reminds of the wordless choirs used to evoke desolate landscapes, such as in Vaughan Williams's *Antarctic Symphony*, or dark and sinister themes, such as in Benjamin Britten's evocation of Auden's "A Summer Night" in his *Spring Symphony*, or, to use a more recent example, John Williams's music for the movie *Harry Potter and the Prisoner of Azkaban*. At once human and *not*-human (because of the rapid vibrato and reverb), choirs of this kind evoke *not*-human landscapes and events while at the same time infusing them with the human emotion of dread.

Clearly, the Roland RD300-SX stands in a long tradition of instruments that on the one hand strive for greater emotive expressiveness and on the other hand use technologies that impel them to reduce human vocal expression to codifiable forms that inevitably strip away its unique "grain" and reduce it to a more abstract idea of "the" human voice. In this, it resembles the paradoxes of contemporary digital images that strive for ever greater

realism, yet can achieve this only by codifying the very phenomena which, in older media, escaped codification and provided the realism of unpredictable randomness and grit. This struggle between the urge to retain a romantic emotivism and realism on the one hand, and the technological imperative of total codification on the other, is likely to continue for some time.

What then constitutes the "human" in this instrument's human voices? Clearly, on the one hand, we have the time-honored qualities of subjectivity, individuality, and emotionality. But on the other hand, humanity has become a surface phenomenon here—a dressing up of a mechanical and relatively character-less foundation. What was once the core, the very vibration of the vocal cords, the very articulation of the vowel sound, now becomes a surface phenomenon, a bit of spicing added to an essentially characterless substance.

A heart no longer beats inside.

Notes

1. Quoted in Fenner Douglass, *The Language of the Classical French Organ—A Musical Tradition Before 1800* (New Haven: Yale University Press, 1969), 95.

2. Ibid., 175.

3. Ibid., 186.

4. Ibid., 210.

5. Ibid., 177.

6. Curt Sachs, *The History of Musical Instruments* (New York: W.W. Norton, 1940), 351.

7. A. Bouman, *Nederland . . . Orgelland* (Leiden: Spruyt, Van Mantgem & De Does, 1964), 100.

8. John Landon, *Behold the Mighty Wurlitzer—The History of the Theatre Pipe Organ* (Westport, CT: Greenwood Press, 1983), 4.

9. Ibid., 28.

10. Sachs, *The History of Musical Instruments*, 415.

11. Peter Williams, *The European Organ 1450–1850* (London: B.T. Batsford, 1966), 288.

12. *RD-300SX Owner's Manual* (Roland Corporation, 2004), 55.

13. Theo van Leeuwen, *Speech, Music, Sound* (Basingstoke, Macmillan, 1999).

14. See, for example, John Lyons, *Semantics Vol. 2* (Cambridge: Cambridge University Press, 1977), and Michael Halliday, *Introduction to Functional Grammar* (London: Arnold, 1985).

15. See Kress and van Leeuwen, *Reading Images—The Grammar of Visual Design* (London: Routledge, 1996).

16. Ibid., 170.

17. Radan Martinec, "Types of Process in Action," *Semiotica* 130, no. 3/4 (2000): 243–268.

18. Sleeve notes of CD of Honegger's *Rugby* and *Pacific 231*, Erato 2292-45242-2. 1985.

19. Roland Barthes, *Image-Music-Text* (London: Fontana, 1977).

imaginary "prelinguistic" universalism through emotive babble, but when it explores ways in which its sonic essence can direct its phenomenological reception. In a sense, this already happens when we audit the predictable beauty in tropes like soaring female arias, innocent children's rhymes, and earthy male ballads: they all "move" through their palpable materialization of mere words.

Yet our aural perception can advance beyond being pleased by mirror images of our idealized selves. In place, we can make contact with the "host of Others" detailed in this essay: pitched breathers, rhythm disruptors, Animist emitters, language resistors, schizophrenic heroes, defeminized screamers, apocalyptic survivors, super-human mannequins, emotional slugs, genealogical declarers, undone rappers, algorithmic speakers, racial robots, and Japanese idols. When the posthuman is vocalized through these figures (and many more besides), the egocentric human in us responsible for reading itself solely in its own image can die a warranted death. Only then can the posthuman be heard.

Contributors

Mark Amerika is an interdisciplinary artist and Professor in the Department of Art and Art History, University of Colorado in Boulder (http://www.markamerika.com).

Isabelle Arvers is a new media curator, critic, and author based in France (http://iarvers.free.fr).

Giselle Beiguelman is a media artist teaching in the Graduate Program in Communication and Semiotics, Catholic University of São Paulo (http://www.desvirtual.com).

Philip Brophy is a musician, composer, sound designer, filmmaker, writer, graphic designer, educator, and academic based in Melbourne (http://www.philipbrophy.com).

Ross Gibson is Professor of Contemporary Art, Sydney College of the Arts, University of Sydney. He makes books, films, and art installations.

Brandon LaBelle is an artist and writer teaching at Bergen National Academy of the Arts, Norway (http://www.errantbodies.org/labelle.html).

Thomas Y. Levin is Associate Professor of German, Princeton University.

Helen Macallan is a filmmaker and research associate in the Creative Practices Group, University of Technology, Sydney.

Virginia Madsen is Lecturer in Media, Music, and Cultural Studies at Macquarie University, Sydney. She is also a radio producer and writer.

Meredith Morse is completing a PhD in art history in the Department of Art History and Film Studies, University of Sydney.

Norie Neumark is Professor of Media Studies, La Trobe University, Melbourne. She is also a sound and media artist (http://www.out-of-sync.com).

Andrew Plain is a film sound designer at Huzzah Sound and filmmaker (http://www.huzzahsound.com/flash_site/flash_site.html).

John Potts is Associate Professor in Media, Music, and Cultural Studies, Macquarie University, Sydney. He is also a radio producer and sound artist.

Theresa M. Senft is Senior Lecturer in Media Studies, University of East London (http://www.terrisenft.net).

Amanda Stewart is a poet, writer, and vocal artist based in Sydney, Australia.

Nermin Saybasili is Assistant Professor, Department of Art History, Mimar Sinan Fine Arts University, Istanbul.

Axel Stockburger is an artist and theorist who lives and works in London and Vienna, where he is a scientific staff member at the Department for Visual Arts and Digital Media/Academy of Fine Arts (http://www.stockburger.co.uk).

Michael Taussig is Professor of Anthropology, Columbia University.

Martin Thomas is Research Fellow, History Department, University of Sydney.

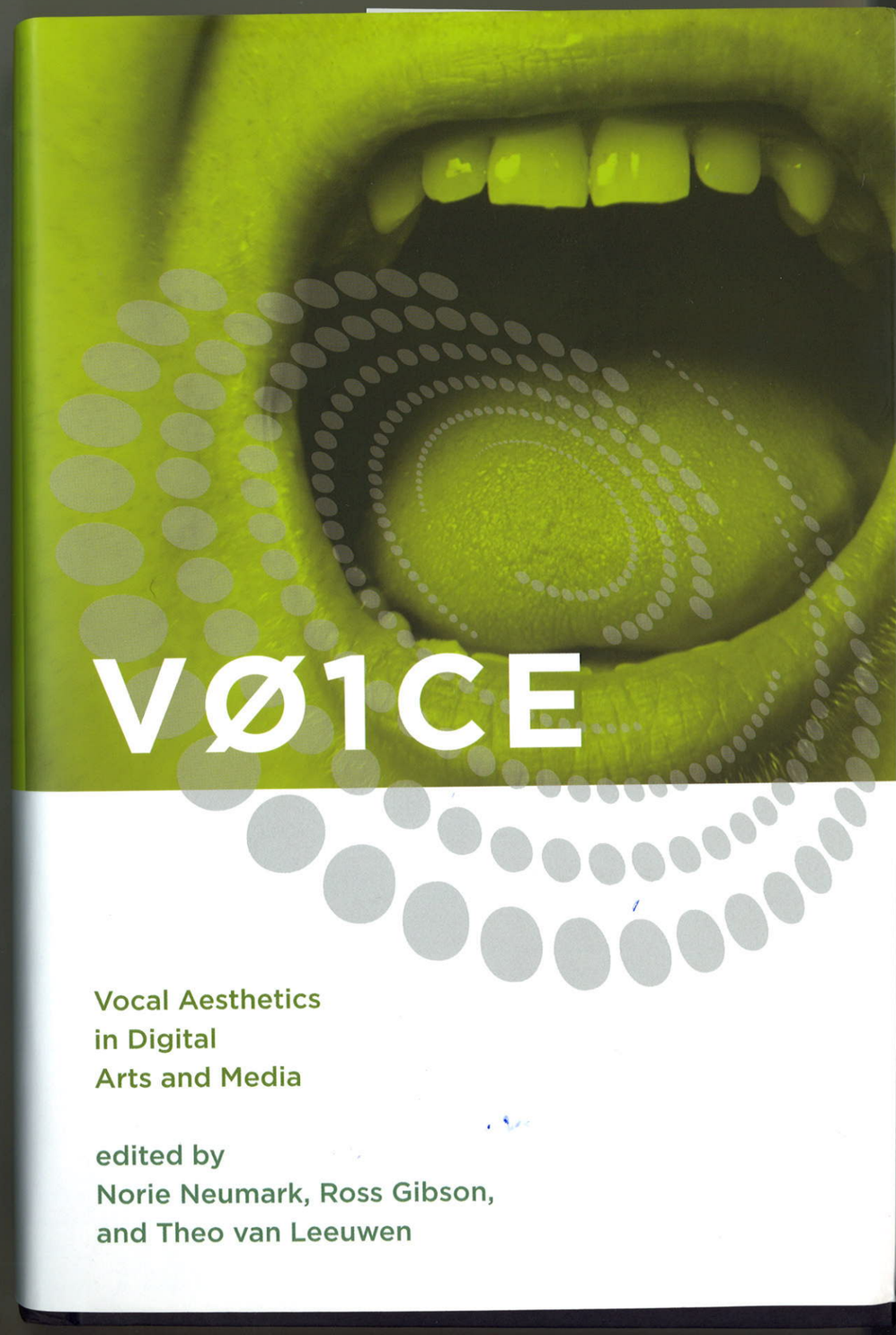
Theo van Leeuwen is Professor of Media and Communication, University of Technology, Sydney.

Mark Ward is Head of the Sound Department at the Australian Film Television and Radio School, Sydney.

Index

Norie Neumark is Professor of Media Studies, La Trobe University, Melbourne. She is a sound and media artist (http://www.out-of-sync.com) and the coeditor (with Annemarie Chandler) of *At a Distance: Precursors to Art and Activism on the Internet* (MIT Press, 2005). Ross Gibson is Professor of Contemporary Art, Sydney College of the Arts, University of Sydney. Theo van Leeuwen is Professor of Media and Communication, University of Technology, Sydney.

A Leonardo Book


VØ1CE
Vocal Aesthetics
in Digital
Arts and Media
edited by
Norie Neumark, Ross Gibson,
and Theo van Leeuwen

new media/technology

Contributors

Mark Amerika, Isabelle Arvers, Giselle Beiguelman, Philip Brophy, Ross Gibson, Brandon LaBelle, Thomas Levin, Helen Macallan, Virginia Madsen, Meredith Morse, Norie Neumark, Andrew Plain, John Potts, Nermin Saybasili, Theresa M. Senft, Amanda Stewart, Axel Stockburger, Michael Taussig, Martin Thomas, Theo van Leeuwen, Mark Ward

The MIT Press
Massachusetts Institute of Technology
Cambridge, Massachusetts 02142
http://mitpress.mit.edu

978-0-262-01390-1

Leonardo

Roger F. Malina, Executive Editor
Sean Cubitt, Editor-in-Chief

Uncanny Networks: Dialogues with the Virtual Intelligentsia, Geert Lovink, 2002
Information Arts: Intersections of Art, Science, and Technology, Stephen Wilson, 2002
Virtual Art: From Illusion to Immersion, Oliver Grau, 2003
Women, Art, and Technology, edited by Judy Malloy, 2003
Protocol: How Control Exists after Decentralization, Alexander R. Galloway, 2004
At a Distance: Precursors to Art and Activism on the Internet, edited by Annmarie Chandler and Norie Neumark, 2005
The Visual Mind II, edited by Michele Emmer, 2005
CODE: Collaborative Ownership and the Digital Economy, edited by Rishab Aiyer Ghosh, 2005
The Global Genome: Biotechnology, Politics, and Culture, Eugene Thacker, 2005
Media Ecologies: Materialist Energies in Art and Technoculture, Matthew Fuller, 2005
New Media Poetics: Contexts, Technotexts, and Theories, edited by Adalaide Morris and Thomas Swiss, 2006
Aesthetic Computing, edited by Paul A. Fishwick, 2006
Digital Performance: A History of New Media in Theater, Dance, Performance Art, and Installation, Steve Dixon, 2006
MediaArtHistories, edited by Oliver Grau, 2006
From Technological to Virtual Art, Frank Popper, 2007
META/DATA: A Digital Poetics, Mark Amerika, 2007
Signs of Life: Bio Art and Beyond, Eduardo Kac, 2007
The Hidden Sense: Synesthesia in Art and Science, Cretien van Campen, 2007
Closer: Performance, Technologies, Phenomenology, Susan Kozel, 2007
Video: The Reflexive Medium, Yvonne Spielmann, 2007
Software Studies: A Lexicon, Matthew Fuller, 2008
Tactical Biopolitics: Theory, Practice, and the Life Sciences, edited by Beatriz da Costa and Kavita Philip, 2008
White Heat and Cold Logic: British Computer Art 1960–1980, edited by Paul Brown, Charlie Gere, Nicholas Lambert, and Catherine Mason, 2008
Curating New Media Art, Beryl Graham and Sarah Cook, 2010
Green Light: Toward an Art of Evolution, George Gessert, 2010
VOICE: Vocal Aesthetics in Digital Arts and Media, edited by Norie Neumark, Ross Gibson, and Theo van Leeuwen, 2010
Enfoldment and Infinity: An Islamic Genealogy of New Media Art, Laura U. Marks, 2010
See <http:mitpress.mit.edu> for a complete list of titles in this series.

VOICE

Vocal Aesthetics in Digital Arts and Media

edited by Norie Neumark, Ross Gibson, and Theo van Leeuwen

The MIT Press
Cambridge, Massachusetts
London, England

For information about special quantity discounts, please e-mail special_sales@mit-press.mit.edu

This book was set in Stone Sans and Stone Serif by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Voice : vocal aesthetics in digital arts and media / edited by Norie Neumark, Ross Gibson, and Theo van Leeuwen.
p. cm. — (Leonardo books)
Includes bibliographical references and index.
ISBN 978-0-262-01390-1 (hardcover : alk. paper)
1. Communication—Social aspects. 2. Voice—Social aspects. 3. Technology and the arts. 4. New media art. I. Neumark, Norie. II. Gibson, Ross, 1956– III. Van Leeuwen, Theo, 1947– IV. Title: Vocal aesthetics in digital arts and media.

P95.54.V65 2010
302.23—dc22

2009038444

10 9 8 7 6 5 4 3 2 1

Contents

Series Foreword

The arts, science, and technology are experiencing a period of profound change. Explosive challenges to the institutions and practices of engineering, art making, and scientific research raise urgent questions of ethics, craft, and care for the planet and its inhabitants. Unforeseen forms of beauty and understanding are possible, but so too are unexpected risks and threats. A newly global connectivity creates new arenas for interaction between science, art, and technology but also creates the preconditions for global crises. The Leonardo Book series, published by the MIT Press, aims to consider these opportunities, changes, and challenges in books that are both timely and of enduring value.

Leonardo books provide a public forum for research and debate; they contribute to the archive of art-science-technology interactions; they contribute to understandings of emergent historical processes; and they point toward future practices in creativity, research, scholarship, and enterprise.

To find more information about Leonardo/ISAST and to order our publications, go to Leonardo Online at <http://lbs.mit.edu/> or e-mail <leonardobooks@mitpress.mit.edu>.

Leonardo/International Society for the Arts, Sciences, and Technology (ISAST)

Leonardo, the International Society for the Arts, Sciences, and Technology, and the affiliated French organization Association Leonardo have two very simple goals:

1. to document and make known the work of artists, researchers, and scholars interested in the ways that the contemporary arts interact with science and technology and
2. to create a forum and meeting places where artists, scientists, and engineers can meet, exchange ideas, and, where appropriate, collaborate.

When the journal *Leonardo* was started some forty years ago, these creative disciplines existed in segregated institutional and social networks, a situation dramatized at that time by the "two cultures" debates initiated by C. P. Snow. Today we live in a different time of cross-disciplinary ferment, collaboration, and intellectual confrontation enabled by new hybrid organizations, new funding sponsors, and the shared tools of computers and the Internet. Above all, new generations of artist-researchers and researcher-artists are now at work individually and in collaborative teams bridging the art, science, and technology disciplines. Perhaps in our lifetime we will see the emergence of "new Leonardos," creative individuals or teams that will not only develop a meaningful art for our times but also drive new agendas in science and stimulate technological innovation that addresses today's human needs.

For more information on the activities of the Leonardo organizations and networks, please visit our Web sites at <http://www.leonardo.info/> and <http://www.olats.org >.

Roger F. Malina
Chair, Leonardo/ISAST
ISAST Governing Board of Directors: Roger Malina, Jeffrey Babcock, Greg Harper, Meredith Tromble, Michael Joaquin Grey, John Hearst, Sonya Rapoport, Beverly Reiser, Christian Simm, Joel Slayton, Tami Spector, Darlene Tong, Stephen Wilson

Preface

There are some questions about the voice and technology that do not go away. What happens, for example, when the distinction between the human voice and the voice of the machine is blurred? Can we still distinguish between genuine and synthetic affect? Should we? Can we still distinguish between the uniqueness of an individual's voice and the social and cultural determinations that shape its performance? Should we even try to do so? This book is an interdisciplinary investigation of questions of this kind as they have been articulated and investigated by historians, philosophers, cultural theorists, film scholars, and artists and designers working in fields such as dance, poetry, music, film, and computer games. The book examines the recording, distribution, and synthesis of the voice, the way voices perform in contexts of technology, and the influence of changing voice technologies on mainstream media such as recorded music, film, and computer games, as well as a range of other themes relating to the ongoing power and centrality of the voice in social and cultural life.

We will focus on digital culture, but our concern is not primarily with the actual way the voice is digitally produced—although that is of interest to a number of the contributing authors. Our primary concern is with the way the voice and its uses are shaped by and help shape digital culture. Our approach therefore avoids technological determinism, attending instead to the complex relations between technology and culture. Many of the chapters in the book demonstrate how the constraints and affordances of digital voice technology were already adopted by artists before digital technology made them seem commonplace or inevitable, thus demonstrating Lewis Mumford's (1939) thesis of the "cultural preparation" that precedes technological innovation, of the way that socially important new technologies