SINGING ABOUT SINGING
Using the Voice as a Tool for Self-Reflection

Rébecca Kleinberger
Bachelor of Engineering
Arts et Métiers ParisTech, 2010

Master of Research in Virtual Environment, Imaging and Visualisation
University College London, 2012

Master of Mechanical Engineering
Arts et Métiers ParisTech, 2013

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, in partial fulfillment of the requirements for the degree of **Master of Science** at the **Massachusetts Institute of Technology**
September 2014

Signature redacted

Author: RÉBECCA KLEINBERGER
Program in Media Arts and Sciences
7 August 2014

Signature redacted

Certified by: TOD MACHOVER
Muriel R. Cooper Professor of Music and Media
Program in Media Arts and Sciences
Thesis Supervisor

Signature redacted

Accepted by: PATTIE MAES
Interim Academic Head
Program in Media Arts and Sciences

SINGING ABOUT SINGING

Using the Voice as a Tool for Self-Reflection

Rébecca Kleinberger

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning on August 8, 2014, in partial fulfillment of the requirements for the degree of **Master of Science** at the **Massachusetts Institute of Technology**

Abstract

Our voice is an important part of our individuality. From the voice of others, we are able to understand a wealth of non-linguistic information, such as identity, social-cultural clues and emotional state. But the relationship we have with our own voice is less obvious. We don't hear it the same way others do, and our brain treats it differently from any other sound we hear. Yet its sonority is highly linked to our body and mind, and is deeply connected with how we are perceived by society and how we see ourselves. This thesis defends the idea that experiences and situations that make us hear, see and touch our voice differently have the potential to help us learn about ourselves in new and creative ways. We present a novel approach for designing self-reflective experiences based on the voice. After defining the theoretical basis, we present four design projects that inform the development of a framework for Self-Reflective Vocal Experiences. The main objective of this work is to provide a new lens for people to look at their voice, and to help people gain introspection and reflection upon their mental and physical state. Beyond this first goal, the methods presented here also have extended applications in the everyday use of technology, such as personalization of media content, gaming and computer-mediated communication. The framework and devices built for this thesis can also find a use in subclinical treatment of depression, tool design for the deaf community, and the design of human-computer interfaces for speech disorder treatment and prosody acquisition.

Thesis Supervisor: TOD MACHOVER
Muriel R. Cooper Professor of Music and Media
Program in Media Arts and Sciences

SINGING ABOUT SINGING

Using the Voice as a Tool for Self-Reflection

Rébecca Kleinberger

The following person served as a reader for this thesis:

Signature redacted

Thesis Reader: RUPAL PATEL

Associate Professor and Lab Director
Department of Speech-Language Pathology and Audiology
College of Computer and Information Science
Northeastern University, Boston

SINGING ABOUT SINGING

Using the Voice as a Tool for Self-Reflection

Rébecca Kleinberger

The following person served as a reader for this thesis:

Thesis Reader: Signature redacted ROSALIND W. PICARD

Professor of Media Arts and Sciences
Program in Media Arts and Sciences
Massachusetts Institute of Technology

Aknowledgements

For their help and inspiration, I am very thankful to:

My advisor: Tod Machover,

The members of my committee: Rupal Patel and Rosalind Picard,

Other professors whose classes inspired me during my two years at the Media Lab: Joseph Paradiso, Neil Gershenfeld, Pattie Maes, Louis D. Braida, John J Rosowski.

The Opera of the Future Group: Ben Bloomberg, Charles Holbrow, Elena Jessop, Taya Leary, Simone Ovsey, Sarah Ryan, Peter Torpey, Akito Van Troyer and Janice Wang.

Collaborators and friends at the MIT Media Lab: Paula Aguilera, Santiago Alfaro, Amna Cavalic-Carreiro, Denise Cheng, Sujoy Kumar Chowdhury, Dhairya Dand, Gershon Dublon, Jorge Duro-Royo, Micah Eckhardt, Mark Feldmeier, Jon Ferguson, Michal Firstenberg, Charles Fracchia, Anette Freiin von Kapri, Pragun Goyal, Elliott Hedman, Javier Hernandes, David Hill, Keira Horowitz, Deepak Jagdish, Alexandra Kahn, Markus Kayser, Oliver Kannape, Steven Keating, Cornelle King, John Klein, Jacqueline Kory, Jared Laucks, Natan Linder, Max Little, Bill Lombardi, Tom Lutz, Brian Mayton, Daniel McDuff, Laia Mogas-Soldevila, Philippa Mothersill, Daniel Novy, David Nunez, Jifei Ou, Will Patrick, Edwina Portocarrero, Linda Peterson, Darthur Petron, Jie Qi, Sandra Richter, Eric Rosenbaum, Akane Sano, Daniel Smilkov, Yoav Sterman, Carlos Gonzalez Uribe, Jonathan Williams and Xiao Xiao,

Friends and moral supports at MIT: Lorin Wilde, Racheal Meager, Anya Burkart, Adi Hollander, Claudio Baroni, Thaddeus Bromstrup, Jindallae Kim, Susan Spilecki, Alan Xu, Thom Howe and Lisa

Chris, Cristina, Sam, Antonio and Nicholas Gold

Jacques Picard, Jacques Bojin, Jacques Paccard, le GP58 et la li209

Marie Weber et Benjamin Lizée,

Pritesh Gandhi and the Ambient Device company

James W Bales and the MIT Edgerton Center,

All the community from La Maison Française in New House,

The bus drivers of Tech Shuttles and Cambridge Saferides West,

Ma family, Mia the tiger, Poopok the dinosaur and Witzy the wolf.

Contents

List of Figures

Prologue / Prelude

In her TED Talk on VocaliD, in 2014, Rupal Patel concluded her presentation with the personal story of a young boy called William who suffers from voice disorder. After several years using the same synthesized voice as thousands of people with the same condition, William was offered, thanks to the VocaliD initiative, a personalized voice created from a donor voice and the few sounds he can himself produce [1]. The first sentence pronounced by William was "Never heard me before. "

"Never heard me before"; after two years working with, deconstructing and rediscovering the human voice, this sentence resonates deeply in very many ways. The specific situation of a boy with vocal disorder let us realize what our voice represents, how precious, personal, and fragile this part of our personality is and how meaningful it can be to listen to one's own voice.

---

[1] Camil Jreige, Rupal Patel, and H Timothy Bunnell. Vocalid: personalizing text-to-speech synthesis for individuals with severe speech impairment. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, pages 259–260. ACM, 2009

1 - Introduction

In the womb, even before our eyes open at month six, our hearing has already been working from the 16th week[2]. In this very isolated environment, all we sense and learn about the outside world as well as our mother's inner state is learned through sounds. The voice[3] and heartbeat of the mother are among the main components of what we hear[4]. As soon as we leave the womb, we are flooded with the complexity of a very different sonic environment. Is there ever a time in our life to reflect on the important calming link that once connected us with ourselves through sound?

1.1 Self-Reflective Vocal Experiences

In this thesis, we define the space of Self-Reflective Vocal Experiences. A Self-Reflective Vocal Experience is a mirror-like installation, situation or experience that is set up to encourage anyone to change their perception and to explore the relationship they have with their voice, and beyond this, the experience they have with themselves.

We all have an intimate experience of our own voice. The voice is infinitely expressive and individually defined. However, many of us do not pay close attention to our voice, do not feel comfortable using it musically or in public, or imagine the use of our voice could be a rich aesthetic experience. It is to address this problem that we are developing a framework to design instances for the exploration of the voice in self-reflective practices.

1.1.1 Self-Reflective

In the *Theaetetus*, Plato describes Socrates 'efforts to define true knowledge. One of the protagonist's queries in this dialogue is to wonder "why should we not calmly and patiently review our own thoughts, and thoroughly examine and see what these appearances in us really are?"[5]

The first step to understanding is to understand oneself, and the first step for this introspection is to be able to turn to oneself, to look at oneself. In our work we don't pretend to give lessons or to point what someone should be looking at. Here we simply propose methods on how to offer

[2] Peter G Hepper, D Scott, and Sara Shahidullah. Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, 11(3):147–153, 1993

[3] Marie Cheour-Luhtanen, Kimmo Alho, Kimmo Sainio, Teemu Rinne, Kalevi Reinikainen, Maija Pohjavuori, Martin Renlund, O Aaltonen, O Eerola, and Risto Näätänen. The ontogenetically earliest discriminative response of the human brain. *Psychophysiology*, 33(4):478–481, 1996

[4] William P Fifer and Christine M Moon. The role of mother's voice in the organization of brain function in the newborn. *Acta Paediatrica*, 83(s397):86–93, 1994; and Lee Salk. The role of the heartbeat in the relations between mother and infant. *Scientific American*, 1973

[5] Francis Macdonald Cornford. *Plato's theory of knowledge: The theaetetus and the sophist*. Courier Dover Publications, 2003

mirror-like situations and experiences to trigger and stimulate the process of understanding the self in a physical, mental and social context.

Even when mirrors intend to show a faithful image of ourselves, they only give us a single perspective. In our work we make the choice of playing on the plurality of perspectives and on the polarizations of the reflected image. If how we see the world is a question of perception, then multiplying the viewpoints when looking at ourselves would give us a rich learning experience.

1.1.2 Vocal

The term vocal is used to qualify a situation or experience in which the voice plays a central role. For encouraging the self-reflection process, the choice of looking at the voice is driven by the multidimensional aspect of the voice itself as a tool, an objective and a bridge.

Voice is a *tool* most of us possess to some extent. It is a *tool* we generally use with great dexterity. It feels very "under-control" and we are used to it. Technologically mediated experience around the voice naturally embeds something reassuring, comforting and human. Because of the apparent familiarity and control we have over our voice, we can use it as a flashlight, to explore its origin, history, mental and physical and social clues and influence. In this context, the voice can be seen as an avatar of the self; they both contain this paradox of being apparently timeless and immutable while being also inherently embedded in the present and shaped by our past experiences (age, traumas, accents, intonation, fillers, vocal posture)[6].

The understanding and broader view of our own voice is also the *objective* of the self-reflective activity. Several instances can be recounted from helping people with disabilities to elaborating new approaches of prosody acquisition for children based more of the understanding of their own body and instrument. Voice as an objective also covers the idea that reflecting on one's own voice can contribute to becoming a more confident speaker, improving affective communication and becoming more perceptive of the emotional values and connotations people convey. Verbalizing is a procedure of cognition and dexterity. This thesis work is based on the idea that we have the option of thought when using our voice and that reflective-type learning can help us gain some mindfulness in this procedure.

Voice is also a *bridge*, because it connects different contexts. It is connected to our mind and body, can it become a connection between those two? Different aspects of the voice have been studied in many distinct research fields: physics, mechanics, signal processing, design,

---

[6] Tobias Riede and Tecumseh Fitch. Vocal tract length and acoustics of vocalization in the domestic dog (canis familiaris). *Journal of Experimental Biology*, 202(20): 2859–2867, 1999

as well as clinical medicine, neurology, biology, surgery, physiology, not forgetting linguistics, philosophy, communication theories, social studies, learning sciences, etc. In order to develop reflective systems, we need to bridge several of those fields to bring an experience that is compelling, comprehensive and enlightening. An experience that brings people in a journey from the inner self to the outside world... and back. In addition, when it comes to our own voice, our perception is greatly altered first because we hear our voice differently than other people hear us, and secondly because when we verbalize, our brain inhibits the auditory cortex. This mechanism can be justified by an optimization standpoint: as we produce the sound, we know what to expect and don't need to spend the energy analyzing it[7]. But between the command of the brain to produce a sound and the final result, there is a great series of physiological, physical and unconscious processes that distort the result from its original target. This timbre of our voice tells a lot about who we are physically and mentally. Maybe this is the voice process that we also want to bridge.

1.1.3 *Experience*

An experience is a situation that extracts us from the default state of mind, requiring the brain to not be in automatic mode. We have an experience when there is something to experience. We have an experience every time we remember that there is something to experience, something to be present for. Experiencing is about existing in the present. All of the efforts presented in this thesis aim to bring the subject back to their own present.

Philippe Manoury describes the "aesthetic quantum" as the "subjectivity mode and primordial conditions for someone to be in a favorable mental disposition to appreciate the aesthetic dimension of what they perceive"[8]. This is how we wish to understand the concept of experience in our work. And why not use the consideration of the aesthetic as a self-learning process?

1.2 *Terminology*

Perceive = per- 'entirely' + capere 'take.'

Exist = ex- 'out' + sistere 'take a stand.'

Experience = ex- 'out' + periri: 'going through'

Strange/Estrangement = extraneus : 'external'

---

[7] John F Houde, Srikantan S Nagarajan, Kensuke Sekihara, and Michael M Merzenich. Modulation of the auditory cortex during speech: an meg study. *Journal of cognitive neuroscience*, 14(8):1125–1138, 2002

[8] Philippe Manoury, Omer Corlaix, and Jean-Guillaume Lebrun. *La musique du temps réel: entretiens avec Omer Corlaix et Jean-Guillaume Lebrun*. Editions MF, 2012

REFLECTIVE INTERACTION is define by Francois Pachet in *The Future of Content is in Ourselves* as "interactions [that] are designed in such a way that users can create objects of interest without being specialists, and through mirror-like, man-machine interactions designed in particular ways"[9]. Because this type of experience design has the potential of being transformative without being normative, it seems interesting to apply it to the voice that is such a personal process deeply and physically connected to our body and mind.

PERSONA from "Per Sonare" ("speaking through") in Latin was the name given to the masks that actors wore during theatrical performances[10]. It was both a way to display their role and an acoustic tool to carry their voice in the theater. Since Carl G. Jung integrated the word in the psychological lexical field, we call persona the part of the personality dealing with the interaction between the individual and the society[11].

VOICE / SPEECH / SINGING: The focus of this thesis is the voice with the verbal content removed. Unless otherwise specified, the modality of the use of the voice (speech, singing, babbling, crying, pure exploration of sounds, etc) is left open. If our background is sometimes specific to one modality, this work aims to group all the voice modalities under one common umbrella. The title of the thesis should be understood more as homage to Seymour Papert, Marvin Minsky and Nicholas Negroponte metacognition theories[12] than as a plea specifically for the activity of singing. Furthermore, some theories consider that, whatever its use, the human voice is always musical, maybe the original music[13].

1.3 Contributions

This thesis develops a new framework for thinking about experiences and interfaces to engage anyone in self-reflection activities through the use of their voice. The contributions from this thesis include:

- A cross-referenced reading of many domains working on voice to extract or present personal elements that the voice reveals. Different aspects of the voice have been studied in many different domains from clinical medicine, biology and surgery to psychology, physics, linguistics, communication theory, learning sciences, etc. However, there are barely any bridges between those different domains. In this thesis, we propose methods and examples on how to use knowledge from those different to make them accessible and experienciable in meaningful way in a self-reflection context.

[9] François Pachet. The Future of Content is in Ourselves. *Csl, Sony Messier, Jeanmarie*, pages 1–20, 2010

[10] R. D. Cherry. Ethos Versus Persona: Self-Representation in Written Discourse. *Written Communication*, 5(3):251–276, July 1988

[11] Carl Gustav Jung. The Analytical Psychology of. pages 1–13, 1953

[12] Seymour Papert. Teaching children thinking. *Programmed Learning and Educational Technology*, 9(5):245–255, 1972; Marvin Minsky. *Society of mind*. Simon and Schuster, 1988; and Nicholas Negroponte. *Soft architecture machines*. MIT press Cambridge, MA., 1975

[13] Siegfried Nadel. The origins of music. *The Musical Quarterly*, 16(4):531–546, 1930

- A history of projects and installations that embrace the characteristics of the Self-Reflection Vocal system. This includes scientific visualization, technical learning about the voice in medical, therapeutic or physical domains; and prior designs of artistic and perceptual experiences.

- Four new systems that embrace the characteristics of the Self-Reflection Vocal framework, including the design concept, interaction model, implementation, and execution of each system.

- A brief outline of future directions, unexplored areas, and example applications for continued research in the domain of guided and unguided self-reflection. Among those application, we expose the potential of these systems in domains such as tools design for the deaf community, design of HCI interfaces for speech disorder treatment or prosody and language acquisition for children, or in the domain of wellbeing and subclinical treatment of depression.

1.4 Thesis Structure

The next chapter: *Background* presents the different fields important for understanding the background works that motivate and support this work. Chapter 3: *Design Studies* starts by covering preliminary explorations that have been at the source of this thesis work. Then this chapter presents four projects that explore ways to engage people individually (or as a group) in participative experiences centered around the voice. All the projects and systems we describe contain certain aspects of the vocal self-reflective framework. Chapter 4: *Conclusion* offers a summary of our work and enumerates possible directions for future research inspired by the theoretical implications of our thesis work.

Throughout this thesis, Design and Art are present "en filigrane" – as in watermark, in the grain of the paper, implicitly, underlying, filigreed along the text – because we think that aesthetic is not a discrete concept but a lens through which one can see the world, therefore those concepts have to be read between the lines.

2 - Background

This chapter puts the thesis in context and presents the different scientific and inspirational domains that impact the development of the Self-Reflective Vocal Experience framework. The first section presents an overview of everyday use of the voice. The second section explores what there is to gain and to learn from listening more thoughtfully. Finally the last section covers prior works in the quest for new modes of expression, channels of communication, and reconnection to the self through sound.

2.1 - Everyday voice

Verbalization is a procedure of cognition and dexterity in which most of us are experts. After some years of learning and acquisition, most of us use our voice every day as a very natural activity. This section will focus on describing this default mode of using the voice by first presenting a succinct understanding of the vocal production process, then focusing on aspects we don't hear about our own voice and yet we perceive about others'. Finally we cover the default mode of human self-perception.

This section does not constitute a compete presentation of the auditory perception mechanism which is outside the scope of this thesis. Only the specific points that enable us to better understand the relationship that people have with their own voice will be developed. For more details on sound perception and the psychology behind it, see Moore and Moore[14].

2.1.1 - Understanding the vocal production process

To understand the basis of voice production and perception, we can follow the flow of the vocal process. We start with the initial mental stage, followed by the physical production of sound.

The initial mental process

The production of vocal sounds starts with cognitive choices followed by nervous activation of the vocal apparatus. In *Speculations on the Control of Speech*, Peter Ladefoged describes speech as a goal-directed activity[15]

[14] Brian CJ Moore and Brian C Moore. *An introduction to the psychology of hearing*, volume 5. Academic press San Diego, 2003; and Stephen Handel. *Listening: An introduction to the perception of auditory events*. The MIT Press, 1993

[15] Peter Ladefoged. A Figure of Speech: A Festschrift for John Laver edited by William J. Hardcastle, Janet Mackenzie Beck. pages 1–14, 1992

"The speech centers in the brain send out instructions for the speech apparatus to achieve certain goals. But it is not at all clear how these goals are defined (...) The major proposal is that we should not consider all aspects of speech to be controlled in the same way. For the most part, consonants are the result of trying to achieve gestural goals such as those specified by articulatory phonology. Tone and intonation are set as auditory goals consisting of pitch patterns. The targets for vowels are points or trajectories in an auditory/acoustic space. The targets for stressed syllables are increased sub-glottal pressure, achieved through actions of the respiratory system."

As an introduction to the mental processes involved in the vocal production, this text gives us a glance at the diversity and the complexity of the sub-processes entailed.

Previous research on speech disorders informs us of the distinction of brain zones corresponding to the mental generation of sentences on one hand and the production of sentences on the other hand. Studies on stuttering[16] show that this condition might be due to too much overlap between these two zones. The study of different types of speech disorders (dysarthria, apraxia of speech, developmental apraxia of speech, developmental stuttering, acquired — neurogenic and psychogenic —, stuttering, and cluttering) helps researchers understand the zones of the brain controlling the voice. We can also tackle the understanding of underlying mental processes by having a look at the field of language acquisition[17] where behavioral studies explore the different stages in child development and the chronology of acquisition of the different vocal and articulatory skills.

The physical production of sounds

The physical process of voice production is very well described by Kenneth N. Stevens in his book on acoustic phonetics — a key reference in the field. He explains all the functioning of the vocal apparatus[18] and the different ways we can physically model the vocal mechanism, using in particular the source-filter model. The descriptions and models bridge many fields from physiology to electronics, psychoacoustics and signal processing. More biological descriptions of vocal functions can be found in the medical research field where voice disorder and laryngology research is a good entry point to understand the regular and non-regular functioning of the vocal apparatus as part of a living organism.

If the mental state often shows through the voice, it is because the physical procedure is closely connected with emotions. Kenneth Stevens

---

[16] Hans-Georg Bosshardt. Cognitive processing load as a determinant of stuttering: summary of a research programme. *Clinical linguistics & phonetics*, 20(5):371–85, July 2006

[17] Maria I Grigos and Rupal Patel. Articulator movement associated with the development of prosodic control in children. *Journal of speech, language, and hearing research : JSLHR*, 50(1):119–30, February 2007

[18] Kenneth N Stevens. *Acoustic phonetics*, volume 30. MIT press, 2000

describes those correlates in terms of vocal modification in situations of strong arousal, where the physical modifications induced will bring about modifications in the voice. For example in the case of anger, fear or excitement, the voice is modified as follow: greater speed and loudness; more energy in high frequencies, extended range of the fundamental frequency (F0), higher median of the F0 due to higher subglottic pressure, disturbed rhythm due to shorter times of speech between two breaths, more precision in the articulation, etc.

In our work a first step of self-reflection is to have participants realize the hidden cognitive and physical complexity of their voice in order to make people aware of the potential enrichment in experiencing their voice differently.

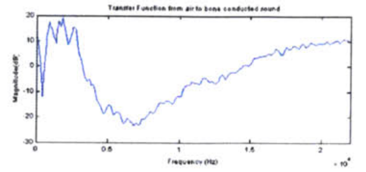


Figure 1: Transfer function of air to bone conducted sound from Berger & al.

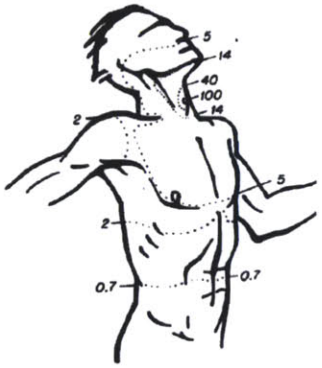


Figure 2: The decline in amplitude of the vibrations from the vocal cords as they travel over the surface of the body. The numbers give relative amplitude from Bekesy & al.

2.1.2 - Perception: What we don't hear but others do

When we say here that *we do not normally listen to our own voice in its everyday use* we mean several things. There are several phenomena (mechanical, neurologic, habituation) that block us to be in contact with our voice.

Physics

We are generally unfamiliar and often uncomfortable when hearing a recording of our own voice. When we talk or sing, we hear a combination of the signal transmitted via air and the one transmitted via cranial bone conduction[19]. The characteristics of this self-perceived voice are a boost of low frequencies and attenuation of high frequencies. The precise transfer function has been studied which enabled the design of systems to simulate the sound of the voice as heard by the speaker[20]. Figure 1 shows the transfer function that applies to self produced sounds as heard by the speaker.

Other internal vibrations can in certain cases reach the ear canal such as noises produced by chewing, swallowing even walking. To optimize the "useful hearing", those sounds as well as our own vocalization sounds are reduced by different mechanisms such as dampening by muscles or by the elasticity of the cervical vertebra, the bones forming the top of the vertebral column[21]. In figure 2 we can see the amplitude dampening of vocal sounds as they spread over the body.

---

[19] Georg V Békésy. The structure of the middle ear and the hearing of one's own voice by bone conduction. *The Journal of the Acoustical Society of America*, 21(3): 217–232, 1949

[20] Jonathan Berger and Song Hui Chon. Simulating the sound of one ' s own singing voice. 2003

[21] Georg V Békésy. The structure of the middle ear and the hearing of one's own voice by bone conduction. *The Journal of the Acoustical Society of America*, 21(3): 217–232, 1949

Brain inhibition

In the same way that the visual cortex has a specific function to recognize faces, the human brain processes voices differently than other sounds[22]. Human audition is specifically tuned to hear voices as suggested by the study of the cocktail party effect. This phenomenon is the human ability to focus our attention on a single speaker when many people are talking at the same time[23]. Beyond this, much previous work indicates that our brain treats our own voice differently than the voices of others. Indeed, even in cases where many different voices in addition to ours are present in the sonic environment, our brain can distinguish between speech sounds that are self-produced and those that stem from the external world. Literature indicates that the auditory cortex is inhibited when processing self-generated compared with played-back speech sounds[24]. By comparing the auditory cortical response to voice in self-produced voiced sounds and in tape-recorded voice, it has been shown that the brain's response to self-produced voiced sounds is weaker[25]. This result suggests that during vocal production, there is an attenuation of the sensitivity of the auditory cortex and that the brain modulates its activity as a function of the expected acoustic feedback. This inhibition has been shown to be present even before the vocal production during mental preparation for speech[26].The brain even prepares itself to not listen to itself.

Feedback loop and reinforcement

It seems obvious that our vocal characteristics do not only depend on our sex and morphology[27] but also on the voices of people around us. Accent and vocal posture carry information about our cultural and social heritage but the characteristics of our own voice also depend on the way we hear our voice, both in specific distribution of our environment and auditory perception, but also in a habituation process of hearing our voice. In other words, our voice is shaped by the vocal environment that surrounds but our voice also shapes this very environment in an habituation process.

The Lombard effect, studied in 1911, predicts that people speak louder when they have difficulties hearing themselves[28]. Since then, many studies show the influence of the auditory perception on the qualities of the voice sound. The auditory feedback influences speech production not only in terms of loudness but also pitch and formant structure[29].

Clues indicate that vocal reinforcement can occur without awareness. In a study conducted in 1998 at the University of Stirling, researchers pretending to work on extrasensory perception asked the subjects to make a guess on which of two cards the experimenter focuses on. The subject was then immediately told if his guess was right or wrong. Among the 160

[22] Marianne Latinus and Pascal Belin. Human voice perception. *Current biology : CB*, 21(4):R143–5, February 2011

[23] Barry Arons. A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 12(7):35–50, 1992

[24] Nadia Müller, Sabine Leske, Thomas Hartmann, Szabolcs Szebényi, and Nathan Weisz. Listen to Yourself: The Medial Prefrontal Cortex Modulates Auditory Alpha Power During Speech Preparation. *Cerebral cortex (New York, N.Y. : 1991)*, June 2014

[25] John F Houde, Srikantan S Nagarajan, Kensuke Sekihara, and Michael M Merzenich. Modulation of the auditory cortex during speech: an meg study. *Journal of cognitive neuroscience*, 14(8):1125–1138, 2002

[26] Mika H Martikainen, Ken-ichi Kaneko, and Riitta Hari. Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex*, 15(3):299–302, 2005

[27] Sarah Evans, Nick Neave, and Delia Wakelin. Relationships between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice. *Biological psychology*, 72(2):160–163, 2006

[28] Harlan Lane and Bernard Tranel. The lombard sign and the role of hearing in speech. *Journal of Speech, Language, and Hearing Research*, 14(4):677–709, 1971

[29] David W Purcell and Kevin G Munhall. Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, 119(4):2288–2297, 2006

trials, the experimenter's assessment only depended on the loudness of the subject's voice (half of the users were rewarded when using a soft voice, the other half when using a loud voice). The results show that without awareness, the exercise results in a modified voice[30].

When it comes to semantic meaning, even in the case of nonverbal vocal sound (foreign language, gibberish, etc,) studies involving functional imaging suggest that an implicit semantic processing takes place in the brain when the subject directs his attention to nonverbal input analysis.[31] Even when no words are present or distinguishable, we process the voice to understand the content.

Some very relevant works in neurology helped to shape our work and thoughts in this thesis. The localization of voice-selective areas in the human auditory cortex in[32] provides the foundation for understanding the neural basis of the "perception of speaker-related features." In[33] the identification of the neural response to speech is shown to be different when the listener is asked to focus his or her attention on the voice or on the verbal content. Studies in auditory habituation both in neurology[34] and biology[35] give us serious support for studying the process of vocal reinforcement: our vocal posture is less due to physical determination than it is due to conditioning.

Those considerations about perception and feedback loop are possible starting points to design Self-Reflective Vocal Experiences. For example, the VOICEminusWORDS project described in section 4.2 is based on the ability from the brain to semantically analyze non verbal inputs from the voice. This project is a real-time signal processing software application that removes the verbal content from spoken sentences while keeping the individuality, expressivity and prosody of the message. The goal is to offer an experience to shift people's attention from the words said to how the words are said.

2.1.3 - Modes of self-reference

Narrative self vs. Experiential self

Since William James[36] proposed categorizations of the notion of self in 1890, his work has been supported, refined and redefined by many researchers establishing links between philosophical perception, cognitive sciences and neurology in the domain of self-awareness. The texts and examples covered in this section distinguish between two important aspects of the self. On the one hand, the "narrative self," also called "extended self-reference linking experiences across time[37]," is defined by S. Gallagher as "A more or less coherent self (or self-image) that is constituted with a past and a future in

---

[30] David A. Lieberman Wendy L. Sunnucks Jamie DJ Kirk. Reinforcement without awareness: I. voice level. *The Quarterly Journal of Experimental Psychology: Section B*, 51(4):301–316, 1998

[31] Katharina von Kriegstein, Evelyn Eger, Andreas Kleinschmidt, and Anne Lise Giraud. Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain research. Cognitive brain research*, 17(1):48–55, June 2003

[32] P Belin, R J Zatorre, P Lafaille, P Ahad, and B Pike. Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309–12, January 2000

[33] Katharina von Kriegstein, Evelyn Eger, Andreas Kleinschmidt, and Anne Lise Giraud. Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain research. Cognitive brain research*, 17(1):48–55, June 2003

[34] B Pfleiderer. Visualization of Auditory Habituation by fMRI. *NeuroImage*, 17(4): 1705–1710, December 2002

[35] D G Sinex and C D Geisler. Auditory-nerve fiber responses to frequency-modulated tones. *Hearing research*, 4(2): 127–48, May 1981

[36] William James. *The principles of psychology*. Digireads. com Publishing, 2011

[37] Norman AS Farb, Zindel V Segal, Helen Mayberg, Jim Bean, Deborah McKeon, Zainab Fatima, and Adam K Anderson. Attending to the present: mindfulness meditation reveals distinct neural modes of self-reference. *Social cognitive and affective neuroscience*, 2(4):313–322, 2007

the various stories that we and others tell about ourselves." On the other hand the "minimal self," also called "experiential self", "phenomenological self" or "momentary self," is defined as "Phenomenologically, that is, in terms of how one experiences it, a consciousness of oneself as an immediate subject of experience, un-extended in time"[38]. After redefining those two selves, S Gallagher shows how philosophical views align with knowledge in cognitive sciences in terms of self-reference quoting as examples self-reference theories from Wittgenstein[39] and Shoemaker[40].

Researchers also established links between neurology and those two modes of self-perception by studying the brain in meditation[41]. The results suggest an essential neural dissociation between the momentary self and the narrative self. The investigators argue that training can help dissociating them like in the case of trained meditation. Neurological studies support the hypothesis of the narrative mode of self-perception as the default mode of self-perception[42].

Interoception/ Introspection

The consciousness of the body and ability to be aware of visceral sensation (gut feeling) is called interoception. It is often measured through a heartbeat timing task, assessing the ability of people to measure their own heartbeat as a measure of their sensitivity toward their internal bodily response[43]. This sensitivity toward our body condition, also called "the material me", has been shown to be correlated with mental self-awareness and ability to access and understand one's own inner feelings[44]. Those findings support the theory presented by Damasio in *Descartes's error*[45] in which our representation of the homeostatic condition of our physical body is the ground of our mental self-representation.

Researchers have identified the right anterior insula (rAI) to be the region of the brain responsible for people's degree of interoception. The activity and size of this brain region correlate with individual differences in the connection people have with their physical feelings[46]. In his study aiming to explain why some humans are more aware than others, Dr. A.D Craig concludes with these words:

"The most important implication of this study emerges from the correlation of both activity and physical size of rAI with individual subjective awareness. This implies that individual differences in subjective interoceptive awareness, and by extension emotional depth and complexity, might be expressed in the degree of expansion of rAI and adjacent orbitofrontal cortices. This clearly resembles the relationship between the size of auditory and motor cortices and the pitch and performance ability of musicians, which is partly innate and partly the result of training"

---

[38] Shaun Gallagher. Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4(1): 14–21, 2000

[39] Ludwig Wittgenstein. The blue and brown books. 1958

[40] Sydney Shoemaker. *Identity, cause, and mind: Philosophical essays*. Oxford University Press, 2003

[41] Norman AS Farb, Zindel V Segal, Helen Mayberg, Jim Bean, Deborah McKeon, Zainab Fatima, and Adam K Anderson. Attending to the present: mindfulness meditation reveals distinct neural modes of self-reference. *Social cognitive and affective neuroscience*, 2(4):313–322, 2007

[42] Malia F Mason, Michael I Norton, John D Van Horn, Daniel M Wegner, Scott T Grafton, and C Neil Macrae. Wandering minds: the default network and stimulus-independent thought. *Science*, 315(5810): 393–395, 2007

[43] Hugo D Critchley, Stefan Wiens, Pia Rotshtein, Arne Öhman, and Raymond J Dolan. Neural systems supporting interoceptive awareness. *Nature neuroscience*, 7(2):189–195, 2004

[44] Arthur D Craig. How do you feel? interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, 3(8):655–666, 2002

[45] Antonio Damasio. *Descartes' error: Emotion, reason and the human brain*. Random House, 2008

[46] AD Craig. Human feelings: why are some more aware than others? *Trends in cognitive sciences*, 8(6):239–241, 2004

This last consideration supports the value of offering people tools for self-reflection based on increasing the awareness of the mental, physical and visceral aspects of the self through the use of the voice.

2.2 - What listening can bring us

From the default mode of listening to the voice – and from that, to the self – described above, we can already foresee the potential benefits of a more thoughtful listening process. The next section shows how using the voice as a flashlight makes it possible to light up and understand physical and health related disorders. A second section covers examples in which the voice and its qualities are the objective and in which better listening and sensing can help enhance the vocal properties. A third part views the voice as a possible bridge within and between people. The fourth part proposes a succinct review of research on analysis and synthesis of the human voice. Finally a fifth part shows links between better listening and richer self-reflection.

2.2.1 - Voice as a tool

Flashlight for Physical health

Bioacoustics is the domain that investigates how living bodies produce, disperse and react to sounds[47]. In this field, different techniques have been developed to extract meaningful sound signals from the body. Monitoring and listening to the sounds produced in the body is a common diagnosis method. The target sound can be either vocal or non vocal. In the case of diagnosis from the recording of voiced sounds, different methods can cover a wide range of diseases.

Respiratory anomalies often modify the sound of the voice because they change the acoustic characteristic of the trachea. Physical modeling tools developed to analyze the human airways as acoustical tubes have previously been used to diagnose the origin of respiratory conditions[48]. In the case of lung diseases, it is the source of the voice and not the filter that is altered. Using time-expanded wave-form analysis, doctors have been able to characterise the subtle discontinuity of speech sounds due to abnormal lung conditions[49]. The origin of chest pain can be associated with many types of physical or mental issues such as cardiac problems, acute coronary syndromes, pulmonary embolisms, pneumonia or panic disorder. In several cases, the analysis of the vocal sounds can help diagnose the origin of the pain[50]. The evaluation of voice dosimetry can help predict the probability of bronchial stenosis on patients who have normal voice sound but weak breath sound[51].

[47] Peter K. McGregor. Bioacoustics journal. URL `http://www.bioacoustics.info/`

[48] René Théophile Hyacinthe Laennec. *De l'auscultation médiate: ou, Traité du diagnostic des maladies des poumons et du coeur; fondé principalement sur ce nouveau moyen d'exploration*, volume 2. Culture et civilisation, 1819; and Amon Cohen and AD Berstein. Acoustic transmission of the respiratory system using speech stimulation. *Biomedical Engineering, IEEE Transactions on*, 38(2):126–132, 1991

[49] Raymond LH Murphy Jr, Stephen K Holford, and William C Knowler. Visual lung-sound characterization by time-expanded wave-form analysis. *New England Journal of Medicine*, 296(17):968–971, 1977

[50] William E Cayley Jr. Diagnosing the cause of chest pain. *Am Fam Physician*, 72 (10):2012–2021, 2005

[51] FL Jones. Poor breath sounds with good voice sounds. a sign of bronchial stenosis. *CHEST Journal*, 93(2):312–313, 1988

Max Little implemented a biologically and mechanically based model that takes into account nonlinearity, non-Gaussian and turbulent aspects present in vocal production[52]. This work has been used in different clinical contexts from the evaluation of speech pathology, breath and lung problems, neuronal malfunctions, or muscular control difficulties, to the detection of early stages of Parkinson's disease. The Parkinson Voice Initiattive enable people to be tested for early stages of Parkinson's deasise with 99% accuracy from a simple phone call[53].

2.2.2 - *Voice as an objective*

Quality of the voice

The voice signal has been studied in sub-clinical context by using sensors in contact with the skin to analyse its vibrations. The anatomic and physiologic characteristic of the nasal canal have been studied in correlation with skin transmitted measurements on the nose[54]. Also the phenomena of perceived nazalisation[55]and hypernazalisation[56]in speech have been studied using the same technique and correlated to the ratio between signal measured at the nose and neck skin.

The use of such monitoring in the case of professional singers[57] reveals the interesting fact that "for some male singers, the amplitude seems to lie just above the threshold for vibratory sensation (...) a male singer can make use of his sensation of chest vibrations as a non-auditory feedback signal for phonatory control". This suggests that learning to become aware of the vibrations could help singers become less influenced by the acoustic characteristic of the halls when assessing their performance. In the same line studies have evaluated the possibility that singers use vibratory feedback unconsciously to enhance tonal control[58].

Vocal disorder

The domain of speech therapy, in cases of motor or neurological speech disorder or for children on the autism spectrum, proposes non-classical methods to help children in the phase of language and prosody acquisition or to help people reconstruct their voice[59]. The use of personalized computerized speech therapy systems by children and getting feedback on their progress from interactive games has shown to have positive effects on the engagement and progress of the participants.[60]

When it comes to experiencing sound, Evelyn Glennie — the first deaf musician to enter the Royal Academy of Music in London — engages her audience during her percussion performances to experience sound with all

---

[52] M. a. Little, P. E. McSharry, I. M. Moroz, and S. J. Roberts. Testing the assumptions of linear prediction analysis in normal vowels. *The Journal of the Acoustical Society of America*, 119(1):549, 2006

[53] Max A Little. *Biomechanically Informed Nonlinear Speech Signal Processing*. PhD thesis, University of Oxford, 2006

[54] Jerry Moon. The influence of nasal patency on accelerometric transduction of nasal bone vibration. *The Cleft Palate-Craniofacial Journal*, 27(3):266–274, 1990

[55] RP Lippmann. Detecting nasalization using a low-cost miniature accelerometer. *J Speech Hear Res*, 24(3):314–317, 1981; and Margaret Redenbaugh. Correspondence between an accelerometric nasal/voice amplitude ratio and listeners' direct magnitude estimations of hypernasality. *Journal of speech and hearing research*, 28(2):273, 1985

[56] Yoshiyuki Horii. An accelerometric measure as a physical correlate of perceived hypernasality in speech. *Journal of speech and hearing research*, 26(3):476, 1983

[57] Johan Sundberg. *Chest vibrations in singers*, volume 22. Centre Georges Pompidou, 1979

[58] Ronald T Verrillo. Vibration sensation in humans. *Music Perception*, pages 281–302, 1992

[59] Adam Boulanger. *Autism, new music technologies and cognition*. PhD thesis, Massachusetts Institute of Technology, 2006; and Rupal Patel, Caroline Niziolek, Kevin Reilly, and Frank H Guenther. Prosodic Adaptations to Pitch Perturbation. 54 (August):1051–1060, 2011

[60] Mohammed E Hoque, Joseph K Lane, Rana El Kaliouby, Matthew Goodwin, and Rosalind W Picard. Exploring speech therapy games with children on the autism spectrum. 2009

their ports, all the channels, opening all the doors and windows and to feel the sound as she does not only with their ears but with their whole body[61].

Tools built for the deaf community can also be an interesting source of inspiration. The Tadoma method, invented in the 20's by Sofia Alcorn, is a communication tool for deaf blind people. Also referred to as "tactile lip-reading", the method is described as follows : "the deaf blind person places their thumb on the speaker's lips and their fingers along the jaw line. The middle three fingers often fall along the speaker's cheeks with the little finger picking up the vibrations of the speaker's throat."[62] This use of alternative senses to be as close as possible to the voice production organ is inspirational because it also brings people closer to the emotion and the aliveness of the voice. Indeed the Tadoma method can be precise enough that it enables people to detect intonation from the throat vibrations[63].

2.2.3 - Voice as a bridge

The use of the voice in musical and traditional practices is very present in every culture. The benefits and values are often linked to social and personal aspects as much as physical aspects. In this way, we can also perceive vocal practices as a means to bridge those three aspects of human life.

Many studies show the benefits of music on people's wellbeing. This topic is outside the scope of this thesis and the following section only presents some specific cases showing the benefits of active vocal practices. For more information about the interaction of music and wellbeing, see[64].

Voice and wellbeing

In the subclinical domain, several studies have focused on the links between singing and the physiological signs of wellbeing (heart rate, blood pressure, and stress level). The World Health Organization Quality of Life (WHOQOL) project is defined by the World Health Organisation[65] as "aiming to develop an international cross-culturally comparable quality of life assessment instrument. It assesses the individual's perceptions in the context of their culture and value systems, and their personal goals, standards and concerns." Using the WHOQOL scale, studies reveal significant benefits among choral singers in England[66]. This combinaison of singing and social activity has been shown to have benefits on four categories of wellbeing challenges: enduring mental health problems; significant family/relationship problems; significant physical health challenges; and recent bereavement. In this study, the main impacts found on wellbeing and health are associated with six "generative machines": positive affect; focused attention; deep breathing; social support; cognitive stimulation; and regular commitment.

[61] By Evelyn Glennie. Hearing Essay. 1993

[62] C M Reed, W M Rabinowitz, N I Durlach, L D Braida, S Conway-Fithian, and M C Schultz. Research on the Tadoma method of speech communication. *The Journal of the Acoustical Society of America*, 77(1): 247–57, January 1985

[63] Hong Zhang Tan, Nathaniel I Durlach, William M Rabinowitz, and Charlotte M Reed. Information transmission with a multi-finger tactual display. *Scandinavian audiology. Supplementum*, 47:24–28, 1996

[64] Raymond MacDonald, Gunter Kreutz, and Laura Mitchell. *Music, health, and wellbeing*. Oxford University Press, 2012

[65] World Health Organization. World health organization. URL `http://www.who.int/substance_abuse/research_tools/whoqolbref/en/`

[66] Stephen Clift, Grenville Hancox, Ian Morrison, Bärbel Hess, Gunter Kreutz, and Don Stewart. Choral singing and psychological wellbeing: Findings from english choirs in a cross-national survey using the whoqol-bref. In *Proceedings of the International Symposium on Performance Science*, pages 201–207, 2007; and Stephen Clift. The significance of choral singing for sustaining psychological wellbeing : findings from a survey of choristers in England , Australia and Germany. pages 79–96, 2010

A study on individual singers reveals that there is a major difference of effects of singing lessons on amateurs and professionals[67]. Both groups had a higher level of oxytocin and reported feeling more energetic after the singing lesson. But only the amateur reported feeling better and using the lessons as "a means of self-actualization and self-expression as a way to release emotional tensions" while the professionals were more achievement oriented.

A very elegant study conducted at the Max Planck Institute for Human Development, with a population of eleven singers and one conductor in a choir, shows that respiratory rhythms as well as cardiac pattern synchronize within and between people during choir singing[68]. This phenomenon is partially explained by heart rate variations due to the difference of blood composition during inhalations and exhalations, also known as respiratory sinus arrhythmia. The same type of phenomenon has been shown to happen during poetry recitation[69].

The study of this synchronization process has been pushed further in a publication linking musical structure and heart rate variability of singers[70]. The authors mention possible implications on joint action and cooperation. Because it has been shown that joint action leads to joint perspectives[71] the researchers propose the possibility that choir singers could have their egocentric perspective on the world modified into a joint common perspective, thus contributing to the explanation of why music and singing have such universal and ancestral roots.

The studies presented above agree on the fundamental importance of breath control, induced by the use of the voice, as an important connection between singing and physiology. Moreover, the social aspect and mindset of people involved in the singing activity is often another fundamental factor to understand the possible health and wellbeing benefits. In this thesis work and especially in the ORB project described in section 4.3, the relf-reflective aspect is at the frontier between the physical wellbeing and the mental meditation.

2.2.4 - Analysis and synthesis of expression and individuality in human voices

Analysis

When it comes to extracting meanings or features from voice, only a few digital signal processing (DSP) tools are really adapted to the specific range, richness and complexity of the human voice. In speech recognition research, for example, the algorithms have to normalize the voice signal,

---

[67] Christina Grape, Maria Sandgren, Lars-Olof Hansson, Mats Ericson, and Töres Theorell. Does singing promote well-being?: An empirical study of professional and amateur singers during a singing lesson. *Integrative physiological and behavioral science : the official journal of the Pavlovian Society*, 38 (1):65–74, 2003

[68] Viktor Müller and Ulman Lindenberger. Cardiac and respiratory patterns synchronize between persons during choir singing. *PloS one*, 6(9):e24893, January 2011

[69] Dirk Cysarz, Dietrich von Bonin, Helmut Lackner, Peter Heusser, Maximilian Moser, and Henrik Bettermann. Oscillations of heart rate and respiration synchronize during poetry recitation. *American Journal of Physiology-Heart and Circulatory Physiology*, 287(2):H579–H587, 2004

[70] Björn Vickhoff, Helge Malmgren, Rickard Å ström, Gunnar Nyberg, Seth-Reino Ekström, Mathias Engwall, Johan Snygg, Michael Nilsson, and Rebecka Jörnsten. Music structure determines heart rate variability of singers. *Frontiers in Psychology*, 4(July):1–16, 2013

[71] Björn Vickhoff. A perspective theory of music perception and emotion. *rapport nr.: Skrifter från musikvetenskap 90*, 2008

therefore removing personal and affective content.

Some of the projects presented in this thesis are based on extraction techniques existing in the general domain of sound signal processing and on the different works already done on extracting prosody features[72] or voice quality elements[73] of affective markers[74] from the vocal signal.

Synthesis

Computers can now reproduce certain important affective markers present in the voice. Work by Janet Cahn[75], on how to generate expression in synthesized voice, offers a good overview of the acoustic and prosodic correlates to emotions. She emphasizes the importance of intonation that is "intertwined with affects", and gives a good summary of the basic voice parameters that are perceptually important to convey expressivity (pitch, timing, voice quality, articulation). On this topic, Raul Fernandez goes further on the mathematical description of voice quality. He highlights the importance of deconvolution of the speech signal into the glottal flow derivative and the vocal tract filter to be able to mathematically access emotional quality of speech[76].

2.2.5 - Better listening: self-reflection and feedback process

In his self-perception theory, Pr. Daryl Bem proposes two postulates:[77]

"When we want to know how a person feels, we look to see how he acts. Accordingly, it looks possible that when an individual himself wants to know how he feels, he may look to see how he acts, as possibly suggested anecdotally by such statements as "I guess I'm hungrier than I first thought." It was from this line of reasoning that the first postulate of self- perception theory was derived: Individuals come to "know" their own attitudes, emotions, and other internal states partially by inferring them from observation of their own overt behaviors and/or the circumstances in which this behavior occurs.

The second postulate of self-perception theory suggests a partial identity between self- and interpersonal perception: To the extent that the internal cues are weak, ambiguous or uninterruptable, the individual is functionally in the same position as an outside observer, an observer who must necessarily rely upon those same external cues to infer the individual's inner state. "

Two considerations taken from those examples are particularly interesting for this work. First, Bem's theory corroborates our hypothesis that our perception of our own behaviors — including our voice — can shape our self-image. On this point, some of the conclusions drawn from the Mood Meter system[78] are very relevant. By simply displaying to you if you smile

---

[72] J. Droppo and a. Acero. Analysis and comparison of two speech feature extraction/compensation algorithms. *IEEE Signal Processing Letters*, 12(6):477–480, June 2005; and Nick Tsakalos and Evangelos Zigouris. Autocorrelation-based pitch determination algorithms for realtime vocoders with the TMS32020/C25. *Microprocessors and Microsystems*, 14(8):511–516, October 1990

[73] Adam P Kestian and Tamara Smyth. Real-Time Estimation of the Vocal Tract Shape for Musical Control. 2010; Youngmoo Edmund Kim. *Singing Voice Analysis / Synthesis*. PhD thesis, MIT, 2003; Y Qi and R E Hillman. Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. *The Journal of the Acoustical Society of America*, 102(1):537–43, July 1997; William David Oliver. The Singing Tree: A Novel Interactive Musical Experience. (1995), 1997; and Rahul Shrivastav and Christine M. Sapienza. Objective measures of breathy voice quality obtained using an auditory model. *The Journal of the Acoustical Society of America*, 114(4):2217, 2003

[74] Christos-Nikolaos Anagnostopoulos, Theodoros Iliou, and Ioannis Giannoukos. Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. *Artificial Intelligence Review*, November 2012; and Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, 31(1): 39–58, January 2009

[75] Janet E Cahn. The generation of a ect in synthesized speech. *Journal of the American Voice I/O Society*, 8:1–19, 1990

[76] Raul Fernandez. *A computational model for the automatic recognition of affect in speech*. PhD thesis, Massachusetts Institute of Technology, 2004

[77] Daryl J Bem. Self-perception theory1. 1973

[78] Javier Hernandez, Mohammed Ehsan Hoque, Will Drevo, and Rosalind W Picard. Mood meter: counting smiles in the wild. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 301–310. ACM, 2012

or not, the system proved to be a "positive mood booster" for the community. Another example in this domains is the HappinessCounter[79] a device that recognizes smiles and sends networked feedback to the user. This device as well showed to improve the mood of the user and improve interpersonal communication in the community.

Bem also mentions several times in his book the notion of reinforcement but generally as a concept that can be manipulated because it takes place in the social sphere. Which brings us to our second consideration and a clarification: In this thesis, the exercises involving perturbation do not aim to manipulate the user's perception with a specific target goal, but they aim to show to the user how relative their judgment and point of view is. The goal is to highlight the gaps and disconnections but leave the initiative of reconstruction to them.

Moreover, though those perception studies tend to show some determinism in the way our perception of the world works, we can maybe argue that, in the domain of voice, we might not have to take for granted the "ambiguous or un-interpretable" aspects of our internal cues, as Bem would consider it. Indeed, one can propose that by learning to pay more attention to some inner cues, people could become more connected to their inner state and actively take charge of their vocal expression.

2.3 - Previous examples of experiences beyond hearing and other modalities to experience Voice

This section presents an overview of experiences beyond hearing. The first part covers a history of projects and installations that present certain aspects of Self-Reflection Vocal system. The second part presents new philosophies of listening both to sound and to the self that leads to new forms of artistic creation.

2.3.1 - Inspiratory Artistic and Perceptual Experiences

Because this thesis explores compelling ways to increase self-perception, several artistic and perceptual experiments became very inspirational, especially when they revolve around the concept of reflective interaction.

BODY TRANSFER ILLUSIONS: The notions of perception and self-perception have been questioned by different studies in the "body transfer illusion" domain with examples such as the rubber hand experiment[80]. In the larger domain of multisensory illusions, we can mention the McGurk effect[81] where the visual and auditory stimuli are incongruous. In this study a person's phoneme production is dubbed with a video of someone pronouncing a

[79] Hitomi Tsujita and Jun Rekimoto. Smiling makes us happier: enhancing positive mood and communication with smile-encouraging digital appliances. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 1–10. ACM, 2011

[80] M Botvinick and J Cohen. Rubber hands 'feel' touch that eyes see. *Nature*, 391 (6669):756, February 1998

[81] John MacDonald Harry McGurk. Hearing Lips and Seeing voices. *1976 Nature Publishing Group*, 1976

different phoneme. The perceived result is often a third different phoneme. In the double flash illusion[82], where a person hears a number of bangs while seeing a different number of flashes, the audio sense proves to be sometimes dominant over the visual one. The study of these types of effect shows the relativity of perception. In everyday situation, humans tend to forget how much each of us is controlled by an instinctive and ancestral brain connection. The brain uses shortcuts to respond faster to situations but these heuristics also allow us to be tricked.

IN THE DESIGN WORLD: and at least since Zeuxis' grapes, creators have always built experiences to play with people's perception of the world. In his visual installations, Olafur Eliasson plays with colored lights to make the world monochrome "because people do not stand in front of his works as if before a picture, but rather inside them, actively engaged, his installations posit the very act of looking as a social experience."[83]. With his work on the Ganzfeld effect[84]in his lighting installations, James Turrell explores ways to change perception through sensory deprivation[85]. Those two examples echo the concept of estrangement and shed a new light on our visual perception of the world. Turrell explains his work with simple words: "for me it was important that people come to value lights", His objective is to awaken a sense of wonder and to remind the audience to not take light for granted even though it is everywhere. It is this same approach that this thesis is bringing to the voice.


Figure 3: Alvin Lucier performing "I am sitting in a room" for the Biennale Musica 2012

"I AM SITTING IN A ROOM" (Alvin Lucier, 1969) is a composition where Lucier explores the interaction between himself and his environment through the interaction between his voice and the acoustics of the room where he sits. In this piece, Lucier is recording himself reading a text and then plays back the recording while recording again. At each step, because of the acoustic characteristic of the room, some frequencies will be amplified more and more while others will be more and more damped. He iterates this process of playing back the most recent recoding while recording it anew until the sound of his voice is completely gone, replaced by the pure resonant frequencies of the room itself. One possible interpretation of this very impressive piece is to remove the "irregularities" that his presence and voice could have created by offering instead the apparition of the particularities of the room independent of him. When the art reveals the physical principles and when the physics reveals some deep aspects of what it means for humans to exist and to project ourselves in our environment, then the top down view of the human condition seems to be impersonated in a very elegant and delicate demonstration.


Figure 4: Setting of Vox Populi

VOX POPULI (Don Ritter, 2005) is an installation in which the user enters an empty room containing a lectern, a microphone and a large

---

[82] Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo. Visual illusion induced by sound. *Brain research. Cognitive brain research*, 14(1):147–52, June 2002

[83] Klaus Biesenbach Rowana Marocci. 2008 Moma Exhibition of Olafur Eliasson work Take your Time, 2001

[84] Walter Cohen. Color-Perception in the Chromatic Ganzfeld. *The American Journal of Psychology*, 71(2):390–394, 2013

[85] James Turrell. Aten Reign (2013),, 2013

video projection of a crowd yelling "Speech! Speech!" to encourage the visitor to get to the microphone and read the prompter displaying historic political speeches. When the user talks, the crowd responds in different ways (laughing, hostility...) depending on the "confidence" of the speaker. The definition of a "confident" speech is a speech with a quick tempo and a high volume. By having a virtual crowd literally urging the user to talk, this installation uses an interesting method to make people use their voice in a way that is not very familiar. The evaluation of the performance uses criteria that are very global while this thesis' work tends to value the personal and individual aspects of the voice.

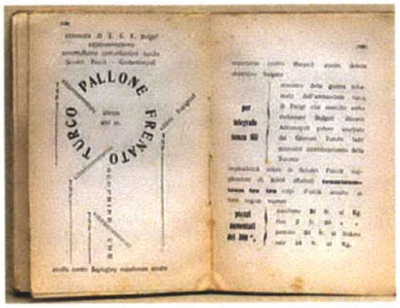


Figure 5: Turkish Balloon, an inner spread from Zang Tumb Tuuum, 1914. This copy is in MOMA's collection.

Sound Poetry: Part of the vanguard genre of concrete poetry, Sound Poetry is an artistic form at the intersection between music and literature based on the phonetic qualities of human voice[86]. Some early examples are found in the Dadaist and futurist movement between the two wars and include Zang Tumb Tumb by F. T. Marinetti or Hugo Ball's performance at Cabaret Voltaire in 1916. The term was invented in 1958 to describe the work of Henry Chopin, largely made possible by technology and the acquisition of a magnetophone. He creates poems of pure sounds. As an example, in "Throat Power," Chopin uses two microphones, one at his lips and one down his throat to create a collage of pulses, gutteralizations, and pounding shreds of sounds. This genre uses the voice and its qualities beyond language to create an art form based on aliveness and performance. Our work and the applied aspects of it, seek to demonstrate to anyone the power and personal control that the voice can vehicle.

Figure 6: Screen shot of Jaap Blonk reciting Ursonate

Ursonography (Golan Levin & Jaap Blonk, 2005) is a multimedia performance of the concrete poetry piece "Ursonate" from Kurt Schwitters, composed to emphasize the abstract musical qualities of speech. During the interpretation, graphics accompany the speech to augment the expressivity in real time. The system also highlights the poem structure through modulations of the motion of the texts in space. The tight connection between the poem itself, the visual augmentation, and Jap Blonk's great performance skills makes this piece very strong and keeps the attention of the spectator very high even though no understandable words are pronounced. The phonetic quality of sound poems in general and the potential in their performability to bring wonder about the richness of human vocal sounds are particularly relevant for our work.

Figure 7: World Premiere of Gesang der Jünglinge, 30 May 1956, Broadcasting Studio of the WDR, Cologne. Used by Permission of Stockhausen Verlag. From (Metzer, 2004)

Stockhausen: Gesang der Jünglinge & Stimmung Written in 1956, Gesang der Jünglinge brings together the voice of a 12-year-old boy and electronics. The composition, highly shaped by mathematical rules and megastructures, unifies the two antipodal sound materials. The composer uses 4 levels of raw vocal material, from a unique sung note to simple

---

[86] Jon Erickson. The Language of Presence : Sound Poetry and Artaud. 14(1):279–290, 2013; and Dick Higgins. U B U W E B :: Dick Higgins on Sound Poetry U B U W E B :: Dick Higgins on Sound Poetry. pages 1–16, 2010

syllables, to words, and then sentences. Following a very complex structural scheme, Stockhausen uses the recording of the child reading a biblical text by applying different treatments inspired by integral serialism[87]. Stimmung is a composition for six singers containing no instruments and no electronics. It is the first important Western composition to be based on the production of vocal harmonics[88]. In each of the 51 sentences, one singer starts the part and the others join to reach "identity" with the material created by the leading singer. The use of overtone singing technique makes the piece highly meditative. The composer describes the piece in those terms: "Stimmung is certainly meditative music. Time is suspended. One listens to the interior of the sound, to the interior of the harmonic spectrum, to the interior of a vowel, to the interior. The most subtle fluctuations—rare outbursts—, all senses are alert and calm. In the beauty of the sensual shines the beauty of the eternal." In the Gesang, Stockhausen creates a sonic continuum from the complexity of the human voice to the simple sine tone. The way the piece is composed strikes the auditory perception: the technology succeeds in creating and highlighting the path to a purification of the voice. In Stimmung, the composition of voices creates a very complex and rich meditative world. The result is access to something both intimate and resulting from the connection of the different parts to reach the unique final "identity" of each of the 51 sentences.

SKINSCAPE (Eric Gunther, 2001) is a tool for composition in the tactile modality[89]. Inspired by knowing whether "the skin [is] capable of understanding and ultimately appreciating complex aesthetic information", this work tackles the relatively uncharted field of tactile composition. The system is based on previous knowledge of tactile illusion, subjective magnitude, adaptation, spatial resolution, psychophysical factors and deduces from those the different parameters important for tactile composition.

Figure 8: Photograph of Skinscape system. (a) V1220 transducer located on wrist; (b) Velcro adjustable elastic band; (c) Interactor cushion amplifier unit; (d) 12-channel V1220 amplifier unit; (e) Digi001 hardware interface; (f) ATX-M3H headphones; (g) Macintosh G4; (h) screenshot of ProTools LE audio editing environment. (From (Gunther, 2001))

SINGING TREEis an artistic installation that is part of the Brain Opera. Brain Opera is a worldwide touring, interactive musical event by Tod Machover seeking to provide to everyone a way to create music[90]. In his thesis[91], William David Olivier describes the Singing Tree as:

> "a novel interactive musical interface which responds to vocal input with real- time aural and visual feedback. A participant interacts with the Singing Tree by singing into a microphone. The participant's voice is analyzed for several characteristic parameters: pitch, noisiness, brightness, volume, and formant frequencies. These parameters are then interpreted and control a music generation engine and a video stream in real-time. This aural and visual feedback is used actively to lead the participant to an established goal, providing a reward- oriented relationship between the sounds one makes and the generated music and video stream one experiences."

---

[87] Robin Maconie. Stockhausen at 70 Through the looking glass. *The Musical Times, Vol. 139, No. 1863 (Summer, 1998), pp. 4-11,* 139(1863):4–11, 2013; David Joel Metzer. The Paths from and to Abstraction in Stockhausen's Gesang der Junglinge. *Modernism/modernity*, 11(4): 695–721, 2004; and John Smalley. Gesang der Jünglinge : History and Analysis. pages 1–13, 2000

[88] Jieun Oh. The Duality of Mechanistic Precision and Timeless Freedom explained in terms of Reaching Identity Goals. pages 1–13, 2009; and Ketty Nez and Hans Tutschku. An Interview with Hans Tutschku. *School of Music 2079 Voxman Music Building University of Iowa Iowa*, pages 14–26, 2003

[89] Eric Gunther. *Skinscape : A Tool for Composition in the Tactile Modality by*. PhD thesis, MIT, 2001

[90] Joseph a. Paradiso. The Brain Opera Technology: New Instruments and Gestural Sensors for Musical Interaction and Performance. *Journal of New Music Research*, 28 (2):130–149, June 1999

[91] William David Oliver. The Singing Tree: A Novel Interactive Musical Experience. (1995), 1997

This work is a good illustration of reflective interaction through voice. By reshaping the way we project ourselves in our environment through our voice, the tree makes the user's voice pass through the looking glass. The effect on the voice is enriched and extended to influence the whole sonic environment in a new and surprising way.

VOICE TRANSFORMATION In Laurie Anderson's work the concept of voice transformation has been pushed in another interesting direction. In the evolution of her art, she found her own way to create "an intensely personal art that is not just simple autobiography"[92]. Coupled with her performance style, she electronically changes the timbre of her voice in order to impersonate several characters. In the same piece and by using different type of filters, she can display different characters, and always proposes dialogues instead of "monolithic" pieces. Close to the notion of personality and research of the self, Laurie Anderson uses voice transformation as a way to let the different characters living in her express themselves.

CHRÉODE is a computer generated tape piece from Jean Baptiste Barrière. The sound synthesis is made using the CHANT program from IRCAM. This program reproduces the qualities of a singing voice using a physically based model of the vocal apparatus. But instead of using a basic filter model, this synthesis makes use of Formant Wave Functions more adapted to the human audible perception. Conceived as an interactive instrument, the CHANT synthesizer is based on physics but controlled by perceptual parameters, such as fundamental frequency, random variations of the pitch, vibrato and random variations, spectrum composition, formants and fundamental, etc[93].

PHONÉ by John Chowning is another interesting example of the exploration of perceptual gradation between two extreme timbres. John Chowning is the creator of the first musical informatics program, built at Stanford in 1964. In 1967, he discovered the value of synthesis by frequency modulation (FM) when applied to electronic music. This method has been very important in that it provides an elegant way to control sound timbre. This technique is also at the core of his piece Phoné, written in 1980. The piece plays on the effect of "bell-changing-to-vowel" using timbral interpolation[94]. In Phoné, Chowning achieves a voice very close to the singing soprano voice. In his notes, he distinguishes six characteristics that helped him in the elaboration of the voice. Those very specific features of the voice are useful for us as perceptual elements. Another input from this work is the using a periodic random vibrato in the perception of vocal tones: "the spectral envelope [alone] does not make a voice!"

---

[92] Mel Gordon. Laurie Anderson, performance Artist. 24(2):51–54, 2013; Jon Mckenzie. Anderson for Dummies. 41(2):30–50, 2009; and Herman Rapaport. " Can You Say Hello ?" States Laurie Anderson ' s United. 38(3):339–354, 2013

[93] Jean-baptiste Barriere. Project : the From Synthesis of to Voice the Singing Synthesis in General The. 8(3):15–31, 2010

[94] John Chowning. "Phoné" (1980-81). 1981; and John Chowning. Fifty Years of Computer Music : Ideas of the Past Speak to the Future. 2007

2.3.2 - Beyond hearing, art under the skin

With development of new technologies flourished new listening philosophies and musical practices. This section describes examples of listening philosophies leading to new creative musical practices

Listening philosophies

One first relevant form of listening philosophy is the concept and practice of Deep Listening[95]. For Pauline Oliveros, the notion of Deep Listening starts from the idea that people are not really listening to what they themselves perform. From those considerations, she began to investigate the processes of human attention and obseve how she herself could affect her emotional state by listening more carefully and with attention. "Prolonged practice brought about a heightened state of awareness that gave me a sense of well-being." From Deep Listening retreats to Deep Listening workshops, Oliveros committed herself to make people understand the value of listening over hearing and considers this practice as a form of meditation in the way it "intends to expand consciousness to the whole space/time continuum of sound/silences."

For composer and revolutionary musician John Cage, the listener's experience of his work is essential to the music itself. His philosophy places silence, noise and sound at the core of music, and the process of listening at the heart of music [96]. Cage's view on noise is very relevant to our work: "Wherever we are, what we hear is mostly noise. When we ignore it, it disturbs us. When we listen to it, we find it fascinating." And for Cage, this fascination does not have to come from a specific interpretation. One does not need to invent a story, or narrative, or life to the sounds outside of what they are. Sounds and music embed a different kind of narrative, their value and life is already in what they are: "They say, 'you mean it's just sounds?' thinking that for something to just be a sound is to be useless, whereas I love sounds just as they are, and I have no need for them to be anything more than what they are. I don't want them to be psychological. I don't want a sound to pretend that it's a bucket or that it's president or that it's in love with another sound. I just want it to be a sound."[97] This view of value of sound in a dimension that is above the narrative meaning, and above the traditional view of aesthetic is very inspirational for our work on voice.

In *Music Mind and Meaning*[98], Marvin Minsky emphasizes the importance of time and human perception of time as one of the main values of music. For him, music teaches us about temporal properties and conceiving objects, concepts and ourselves in time the same way as motion and games helps us learn about space. "But how on earth does one learn about time?

---

[95] Pauline Oliveros. *Deep listening: a composer's sound practice.* iUniverse, 2005

[96] John Cage. The future of music: Credo. *Silence: Lectures and Writings by John Cage,* pages 3–7, 1973

[97] Interview in documentary "listen", 1992

[98] Marvin Minsky. *Music, mind, and meaning.* Springer, 1982

Can one time fit inside another? Can two of them go side by side? In music, we find out!" Music is the present but a present that contains rhythms, redundancies, repetitions of the past and expectations of the future. Even though Minsky specifies that his theory is based on the appreciation of "good music," we can envision the value of time perception when applied to specific listening tasks.

Another paramount legacy from Minsky can be found in his book *Society of Mind* in which is described a model of distributed cognition that explains human intelligence as constructed by the interactions of simple units called "agents."[99] This becomes especially relevant for us when Minsky illustrates his music/time-perception theory through his distributed cognition model: The sound listening philosophy can be understood here as a self-listening paradigm.

Leading to new creative musical practices

By composing Brain Opera, an interactive musical installation based on Marvin Minsky's 'society of mind' theory, Tod Machover and Joe Paradiso instituted a new model for museum experience and for the process of musical creation[100]. Brain Opera is a large interactive multimedia exhibition designed in three phases. In the first phase, visitors enter the Mind Forest[101] where they engage in musical creation via a large set of musical controllers. Gesture Wall, Sensor Chair[102], Rhythm Tree[103], Harmonic Driving, Singing Tree[104], Melody Easel and Digital Baton[105] offer many new ways for the public to interact with sounds, melodies, rhythms and tonalities. Some of those instruments are designed for solo experiences, while others fit to large-scale interactions.

The second phase of the installation is an actual performance from professional musicians. This performance incorporates and makes comprehensive the musical material created by the audience in the previous chamber. The Brain Opera installation can be seen as one of the first attempts to answer the concern raised by Tod Machover: " We seem to want the arts around us all the time, but do not seem to want to let them under our skin. Much remains to be done to bring the public beyond surface titillation and to truly teach the magic of creative expression."[106]

This novel approach to musical creation helps amateurs take part in musical creation, but it is also intended to provide a perspective of the balance between anarchy and central organization within and between the participants through artistic creation. Using the music both as a means and as an objective to reflect on the dichotomy of consciousness as a diverse unity, this approach is particularly relevant for us in terms of plurality of

[99] Marvin Minsky. *Society of mind*. Simon and Schuster, 1988

[100] Joseph a. Paradiso. The Brain Opera Technology: New Instruments and Gestural Sensors for Musical Interaction and Performance. *Journal of New Music Research*, 28 (2):130–149, June 1999

[101]

[102] Tod Machover. Brain opera: Sensor chair, a. URL `http://web.media.mit.edu/~joep/SpectrumWeb/captions/Chair.html`

[103] J Paradiso, J Smith, M Reynolds, M Orth, R Ciliberto, J Rosenberg, M Bach, B Denckla, P Pelletier, and Knaian. Brain opera: Rhythm tree. URL `http://www.media.mit.edu/~joep/TTT.BO/tree.html`

[104] William David Oliver. The Singing Tree: A Novel Interactive Musical Experience. (1995), 1997

[106] Tod Machover. Future arts: By and for whom, and for what? *IBM systems journal*, 39(3.4):678–679, 2000

the self and reflective experiences.

The question of the importance of presence and present during a performance has been tackled in several other projects by the Opera of the Future group at the Media Lab. In Death and the Powers composed by Tod Machover and libretto from Robert Pinsky, the main character leaves the stage during the first scene after downloading himself into what he calls the system. During the rest of the opera he is not physically present on stage but controls the dramatic set using a disembodied performance system to be alive through objects and displays. This very meta view of a performance challenges what it means to attempt a live show. If a performance contains liveness, it makes the audience feel alive, or be alive in the present moment. No matter if the content is virtual or real or recorded or improvised. What matters is if it reaches the objective of existence – making the audience "exist" literally taking a stand outside of themselves. Maybe a way to come closer to this objective is to make people understand, each time in a personal way, that this momentary experience is an experience of selfhood.

3 - Design Studies

In this chapter, we present preliminary explorations followed by four different designs of Vocal Self-Reflective Experiences, all of which share underlying technologies and research developments. From the research involved in their development and deployment, we extract a set of design principles.

3.1 - Preliminary explorations

This section presents six preliminary projects developed during the author's enrolment at the MIT Media Lab. These projects are Vocal Signature, High Speed Vocalization, Vocal Entropy, RoboticoProsody, MiniDuck and Vocal Monitoring through Vibrations. Each of these projects studies the human voice from a different angle, appealing to different domains including information theory, mechanics, social interaction, semantic and phonology.The explorations presented in this section bring us specific findings that give us insights for the design of the four main projects presented in the rest of the chapter. In a broader sense, the collection of these investigations informs how we can build technologies to learn about ourselves through our voice.

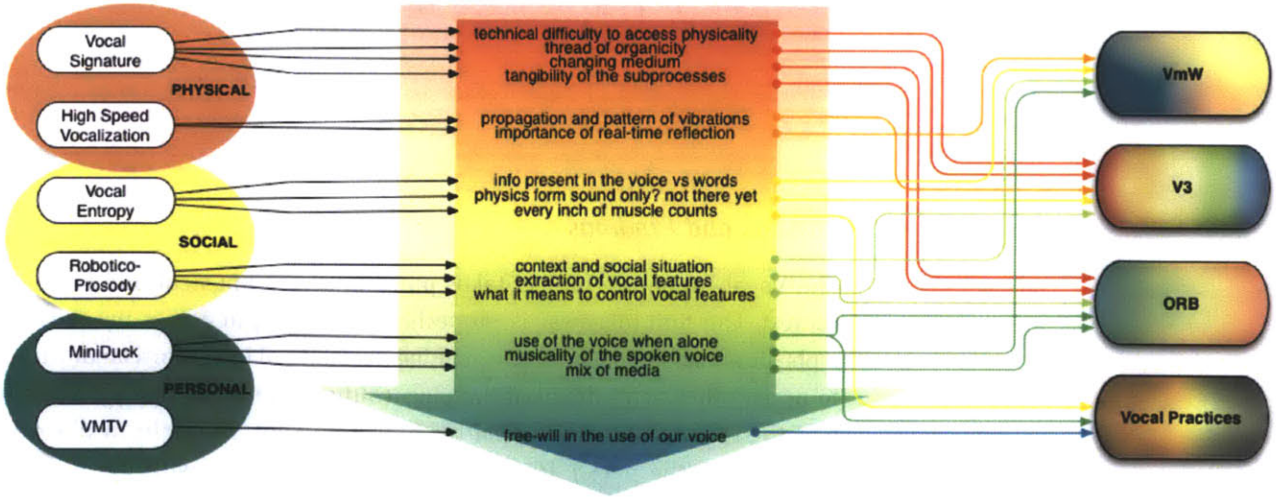


Figure 9: Graphics informing how the preliminary projects inform the design of the four main projects

Figure 9 tells the story of how the findings of each preliminary project map and inform the design of the four main projects. The parallel between voice and self enables us to understand the necessary steps to include in the designs of Self-Reflective Vocal Experiences. Firstly, the voice arises as a physical and embodied experience; then, as it is diffused and perceived by others, it participates in the human social experience (this sharing can be reflected back to us by perceiving how our voice transforms our social and physical environment); finally, a last type of self-reflection is the one resulting from a more personal introversion process.

3.1.1 - Vocal Signature

Project

This project is a calligraphic representation of a voice. The result is a visualization of the physical and mechanical process of vocal production. The contractions and motions of the muscles are analogically projected to a 2D surface. A small mirror is attached to the active parts of the user's neck and the small rotations and motions of the mirror control the trajectory of a laser ray pointing at the mirror. A camera captures the trajectory of the laser point and the video is then post-processed to obtain an effective visualization of the verbalization (Figure 10.) The variations in the result reveal elements of the person's affective state, as it shows through in the muscular movement. Figures 11 and 12 show examples of static visual results.

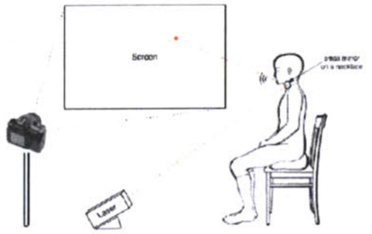


Figure 10: Vocal Signature setting

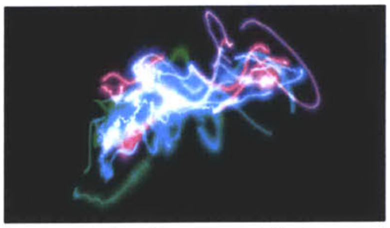


Figure 11: Vocal signature of Haiku

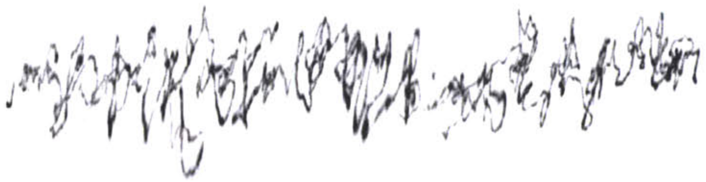


Figure 12: Vocal signature of Verlaine's *Mon Reve Familier*

Lessons and Findings

The Vocal Signature project takes place in a very physical and tangible context. Our first takeaway is the technical difficulty to access the voice as a physical embodied mechanism and the necessity of using "artisanal tricks" to project this small detail of the physicality of the voice (the participant had to sit very still and the mirror hat to be kept precisely in place in contact with the neck.) But the use of a tangible link with the body to access the voice is quite important in our vision.

This is indeed our second takeaway from this work: several previous works have created interactive visual installations with the voice that look very alive and organic. But often, those systems extract continuous parameter from the audio but use physical modeling techniques in the visualization part (particle systems, elastic spring meshes, fluid simulations or cloud simulations). In our work, though the design aspect is less present, the liveness is not added afterword in the process; rather, we intend to not break the thread of organicity coming from the voice. In the V3 project (see Section 3.3) we use and push the idea further by accessing the physical vibrations of the voice from different locations of the face.

A third learning from this project is the potential of changing medium when considering the voice. From our visualizations, we have temporal screenshots of sentences and we can easily compare the same sentence said with different intonations or said by different people. We developed further this idea of changing medium to access the voice in the V3 project (see Section 3.3) and in the development of the ORB (see Section 3.4.)

Another lesson learned here is the idea that lots of what we project through our voice in everyday use is only analyzed by the subconscious part of our brain. Many sub-processes involved in the voice are likely to transmit meta-information about the speaker's mood or mental state. Any muscle can betray us or but it also means small details can help us reveal our mental state to ourselves. Different aspects of the ORB project (see Section 3.4) reuse this idea.

3.1.2 - High Speed Vocalization

Project

The vocal apparatus is mainly hidden from sight, and the voice as a sound is invisible. In the High Speed Vocalization project we stretch time and space to access the very subtle high-speed vibrations present on the skin of the face and invisible to the naked eye. We stretch time by using a high-speed camera shooting up to 1000 images per seconds. We stretch space by magnifying the subtle motions of the skin. The output is a video clip showing visibly the vibrations from the voice propagating in the body.

Different domains of prior work inspired this project. Some recent work has proposed methods to extract audio information form flowing dust captured with a high-speed camera [107]. The location of up to two audio sources can be retrieved from the images captured by a pair of high-speed cameras. Some prior work has been done on the possible extraction of voice information using a high-speed camera [108], but those techniques do not offer comprehensive visualization and require the use of a laser

[107] Mariko Akutsu and Yasuhiro Oikawa. Extraction of sound field information from flowing dust captured with high-speed camera. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 545–548. IEEE, 2012

[108] Mariko Akutsu, Yasuhiro Oikawa, and Yoshio Yamasaki. Extract voice information using high-speed camera. 19:055019–055019, 2013

Doppler vibrometer to access parameters such as the main frequency of the voice. Those technologies and investigations are argued to have potential applications for surveillance and undercover work. The use of high-speed cameras in a more intrusive manner to film the motion of the vocal fold has enabled speech researchers to refine their knowledge in the precise motion mechanism involved. The vocal apparatus being hidden from sight, high-speed imaging of the vocal folds with synchronous audio recording has enabled refinement of the source model to a more accurate one [109].In the domain of speech recognition, the integration of automatic lip-reading algorithms working from visuals, enable better recognition results [110]. But lip-reading supposes that we know the language spoken by the person and that the sounds produced respect certain rules. Our system aims to access something more universal in the voice. In our case, the voice is considered in a more open-ended and organic way: we do not search for patterns, we only search for vibrations.

Implementation

To stretch time, we used a high-speed camera borrowed from the MIT Edgerton Center (Phantom military model.) The participant has once again to sit still on a chair and to produce vocal sounds. The setup includes this camera pointing from various angles to a participant's neck or face and a lamp pointing perpendicularly to the same point. When it comes to camera settings, the rate is chosen to be higher than twice the highest fundamental frequency of the participant's range in order to respect Nyquist's law. The camera sensitivity is chosen to maximize the contrast between the line of the skin and the background. Once the video is recorded, several stages are necessary to process the result. The system used for stretching space has been developed by researchers at MIT CSAIL and is based on spatial decomposition and temporal filtering of the video to amplify small hidden motions at a particular frequency [111]. The result we obtain is a slow motion video showing a spatial accentuation of the skin vibrating during voiced sounds.

Lessons and Findings

Once again the project offers a vision of the voice as a very physical embodied mechanism and reflects to the participant an image of the self and the voice as possible instruments of exploration. The main insight from this project is to see how much the vocal vibrations propagate in the body. The location and pattern of those vibrations has the potential to give clues on the state of the physical vocal apparatus. This project informed the design of the Hexauscultation device presented in Section 3.3. The V3 project pushes this idea further by visualizing in real time those vibrations.

---

[109] Yen-Liang Shue and Abeer Alwan. A new voice source model based on high-speed imaging and its application to voice source estimation. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 5134–5137. IEEE, 2010

[110] Paul Duchnowski, Uwe Meier, and Alex Waibel. See me, hear me: integrating automatic speech recognition and lip-reading. In *ICSLP*, volume 94, pages 547–550. Citeseer, 1994

[111] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics (TOG)*, 31(4):65, 2012

A second thought emerging from this project is the importance of the real-time aspect. Indeed, in this project, the result of the video is very interesting and can help us learn a lot but the downside is that it can not be seen in real time which makes it harder for the participant to connect to what he sees. For the rest of our project and in the design of Self-Reflective Vocal Experiences, we emphasize the idea that the reflection has to be instantaneous for people to really interact with the system. In addition it seems that interacting with a real-time system makes people feel less embarrassed when exploring themselves.

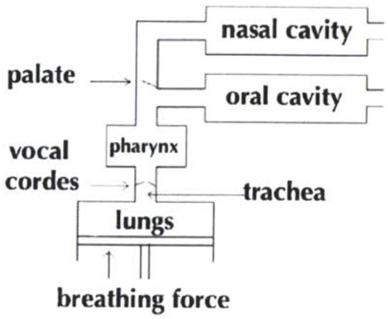


Figure 13: Schematic of the fonctions of the human phonatory aparatus. This modelisation is the one used in our physical model

3.1.3 - Vocal Entropy

Project

The vocal entropy project is a numerical computation based on the understanding of the physical processes of the human voice-box and on the application of governing equations and mechanisms of physics. The project uses information theory to establish a proof of concept to show the bit-saving potential of considering the physical and biological aspects of vocal production when digitally transmitting vocal information. From the understanding of the physical phenomena involved in the vocal production, we built a physical model that reproduces the voice source filter model based on the physical mechanisms involved. Our system can output the sound of different vowels as pronounced by a man's voice. More details are available on the project website[112]. We compute the amount of information in a vocal signal from a physics standpoint. This amount of information is then shown to be much smaller than in a random audio signal: because of how it is produced, the vocal signal offers the potential to be encoded in a much more efficient way.

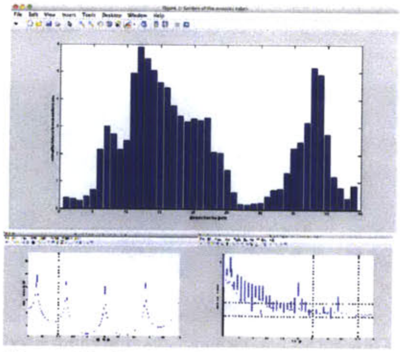


Figure 14: Results from our model of reconstruction of a U sound from the physical shape of the vocal tract. The visualisation on the top represents the sectiosn of the acoustic tubes simulatinghte vocal tract shape, the image on the bottom right represents the resulting formants shape, the figure on the right represents the resulting frequency response

Lessons and Findings

The consideration and learnings from Vocal Entropy concern both the voice as a physical instrument but also its use in the social space of interpersonal vocal interaction. This project was inspirational for the rest of the design projects presented in this thesis because it offered a clear way to see that every inch of muscle counts for the vocal result as well as showing why and how. This idea is explored in the project called "Measurement of Vocal Practices" in which we establish links between people's vibration patterns during vocal practices and their physiological state (see Section 3.5).

The computation presented in this project also gives a glimpse of the difference of scale between the entropy present only in the words and the entropy present when a human voice pronounces the words. This gives us a mathematical support to affirm that in a vocal message, the largest part

[112] Rebecca Kleinberger. Vocal entropy project webpage. URL `http://fab.cba.mit.edu/classes/862.13/students/rebecca/`

of the information is not in the words but really in the voice itself. The VmW project (Section 3.2) proposes a way to bring awareness to this.

A last useful comment comes from the bibliographical research done here on how to reverse-engineer the vocal sound signal to extract data about the physical source. Our project emphasizes the fact that despite the great research done in this domain, scientists are still not able to extract all the information from the vocal sound. This thought was inspirational for the design of the V3 project (Section 3.3) and our last project, Measurement of Vocal Practices (Section 3.5)

3.1.4 - RoboticoProsody

Project

This project is a collaborative work with Jacqueline Kory and explores the way people express themselves vocally when they communicate with technology. We include this study in the Preliminary Exploration part of this thesis because the results give an insight into the voice when used in a social context. The subjects were interacting verbally with a robot. For half the participants, the experimenter introduced the robot in a social way; for the other half, the robot was introduced as a machine. Parameters from the vocal interaction were extracted to analyze the degree of verbal mimicry occurring in the conversation. Details on this project can be found in the publication that resulted from it in HRI 2014 [113].

Lessons and Findings

Figure 15: Two DragonBots, the robot used in this study

As in Vocal Entropy, this project focuses more on the social aspects of the use of voice. Our results showed that the way we frame the robot has an effect on people's behavior: people who perceived the robot more socially spoke significantly more and were more vocally expressive than participants who perceived the robot as a machine. In addition, the range of intonation and vocal mimicry was higher when the participant had been prompted to perceive the robot as a social agent. This is particularly interesting for our work because communicating vocally with technology is still very un-natural for many of us. Indeed, when we use a voice recognition system such as Siri, the use of our voice is very different from when talking to another human being. The importance of social context and use of the voice out of the social context were important in the design of the VmW project and informs the Measurement of Vocal Practices project (Sections 3.2 and 3.5.)

Secondly, another takeaway concerns the techniques used to analyze the participant's voice. To measure engagement and mimicry we tried several

---

[113] Jacqueline Kory and Rebecca Kleinberger. Social agent or machine? the framing of a robot affects people's interactions and expressivity

tools designed for regular sound processing. This project helped us realize how much the vocal sounds behave differently than other sounds, for example, pitch extraction from voice is a particularly complex process because not all sections of speech are voiced and also because the fundamental frequency is mathematically defined as the most preeminent formants but the perceived F0 is sometimes different from the formant with maximum energy[114]. Those learnings in the domain of speech feature extraction were very interesting in the earlier stages of the ORB implementation (Section 3.4). Also, the analysis of the vocal performances in Section 3.5 was inspired by this work.

Another side result from this experience comes from the way we designed the interaction. Two experimenters were needed for this study. One experimenter was welcoming the participants and in charge of the framing, the other one was operating the robot from a different room. We manipulated how the robot was framed at the beginning of the interaction. The study had two conditions. In the Social condition, the experimenter introduced the robot in a social way, by addressing the robot directly and using the second-person "you" to talk to the robot. In the Machine condition, the experimenter introduced the robot as a machine, referring to the robot in the third person and talking about it rather than to it. The robot operator was blind to the condition. Participants were randomly assigned to a condition. To make the interaction more engaging we decided to control the robot in real time which means the operator had to carefully control her vocal features to keep the experience the same for each participant. This idea of controlling vocal features appeared to be more complex than expected. Many rehearsals sessions were necessary to obtain a satisfactory result. This work was close to theatrical rehearsing and reminded us how difficult it is to be aware of the physical details to obtain the chosen social results. This V3 project (section 3.3) offers a way to reflect on some of those vocal features while the last project described in this thesis investigates how those features not only act on the people who hear us but also on our own inner organism.

3.1.5 - MiniDuck

Project

MiniDuck is a collaborative project between Akito Van Troyer and the author that was presented as part of "the Other Festival." It was exhibited on Saturday 20th April among 50 other projects for the *Art and Design Festival* at the Media Lab [115]. This project has the outside appearance of a normal book but is sonically interactive. Each uniquely designed page tells a story about sounds and music and contains text and graphical elements inspired by visual poetry and the Oulipo poets. We used different

---

[114] Jan Bartošek Václav Hanžl. Comparing pitch detection algorithms for voice applications

[115] Tod Machover. The other festival, media lab arts and design festival, b. URL `http://www.media.mit.edu/events/other/projects`

graphic techniques to engage the reader to touch and connect with the drawing to enrich his intellectual experience with touch and music. When the user touches the text on the page, the book plays music corresponding to the vocal musicality of the text. The music is composed from recordings of someone reading the text. We then extracted the pitch contour and harmonized the score. Then we chose a timber associated to the semantic of the page. The resulting MIDI score was associated with the page. The system recognizes which page is currently open and detects which word or group of words the finger is pointing at and plays the according music. MiniDuck translates the musicality of the reading of a book from a textual medium to a musical medium and aims to make people aware of the omnipresence of music in their everyday life. While the reader indulges in reading the book, the specially processed audio accompanies the reader through the reading and the semantic understanding of the texts. More details are available on the project website[116].

Lessons and Findings

After physical or social types of reflection, this project offers a more intimate, personal and traditional way to learn about ourselves: through the action of reading. As a pure mental process, the action of reading acts on our imagination and our perception of the world.

An important insight of this work is the focus on a vocal activity one does alone, using our voice only for ourselves and non defined by social needs or conventions. The reading of a book even silently activates voice related areas in the reader's brain[117]. People generally hear their own "inner voice" pronouncing words mentally. We generalize this idea of an inner-centred use of the voice in several of the main projects by creating activities engaging the public in solo vocal practices in order to interact with the systems. We believe that such a use of the voice already itself has self-reflection virtue.

In this project, the music played by the book at a specific page is composed using the musicality of everyday speech. In the VmW project (section 3.2) we reuse this link between voice, speech and music by removing the words from speech to let people appreciate what remains.

Finally, after having explored projects in which we change the medium from audio to something else, this project mixes three media into one experience: sound, sight and touch. The multiplicity of media does not aim to create a whole new experience but to emphasize what is already present in the normal traditional experience of reading[118]. This idea of trying to unveil existing latent human experience is one of the commun threads of

---

[116] Kleinberger Rebecca Van Troyer Akito. Miniduck webpage. URL `http://web.media.mit.edu/~akito/Projects/MiniDuck/`

[117] Marcela Perrone-Bertolotti, Jan Kujala, Juan R Vidal, Carlos M Hamame, Tomas Ossandon, Olivier Bertrand, Lorella Minotti, Philippe Kahane, Karim Jerbi, and Jean-Philippe Lachaux. How silent is silent reading? intracerebral evidence for top-down activation of temporal voice areas during reading. *The Journal of Neuroscience*, 32 (49):17554–17562, 2012

[118] WD Gaillard, M Pugliese, CB Grandin, SH Braniecki, P Kondapaneni, K Hunter, B Xu, JR Petrella, L Balsamo, and G Basso. Cortical localization of reading in normal children an fmri language study. *Neurology*, 57(1):47–54, 2001; George A Ojemann. Individual variability in cortical localization of language. *Journal of neurosurgery*, 50 (2):164–169, 1979; and Julio González, Alfonso Barros-Loscertales, Friedemann Pulvermüller, Vanessa Meseguer, Ana Sanjuán, Vicente Belloch, and César Ávila. Reading< i> cinnamon</i> activates olfactory brain regions. *Neuroimage*, 32(2):906–912, 2006

this thesis work.

3.1.6 - Vocal Monitoring through Vibrations

Project

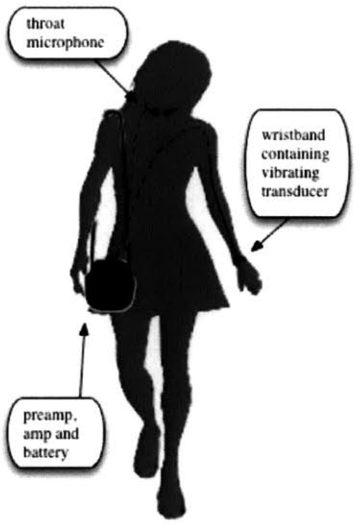


Figure 16: Schematic description of the the device for vocal monitoring through vibrations

This project is a portable wearable device to help people self-monitor the use of their voice. It is constituted of a throat microphone, a preamp, an amp fed by a battery and a vibrating transducer enclosed in a wristband. The whole system is analogic and fits in a small bag or a large pocket. The goal was to see the possible awareness one could gain from monitoring one's vocal activity throughout the day. Many new technologies revolve around the concept of *data-centric life management*. This project proposes to help people manage and have easy access to way they use their voice. By changing medium and offering an element of estrangement, this device can help people notice certain trends in the use of their voice. For instance, people can notice specific personal elements such as: they talk louder outside than inside buildings; they didn't pronounce a word during a group meeting; they are monopolizing the sound space when chatting with a shy friend; they talk with a higher voice when they are on the phone, etc. Only a first prototype has been developed so far as a proof of concept and we think that a next iteration should incorporate more elements of personalization and also learn and adapt to the speaker. A future system should also be able to collect numerical data about the use of the voice. People with voice disorders could also use such a device to gain better control of specific vocal features.

Lessons and Findings

The main takeaway from this project is to question the free will behind the use of voice. So much determinism controls how and when we use our voice that we want to believe in the freeing values of practices for using the voice in a more self-centered way. Music and singing in a way have this potential especially in improvisational situations but we also believe in possible intermediary practices to help those of us who don't yet feel confortable using our voice simply for our own pleasure. To help those people slowly climb the steps enabling them to feel more connected to their voice, to help people sing to themselves inner lullabies. Our voice is a gift we give to other people. How can we use it as a gift to ourselves? We believe that if we learn to cherish it more, it would then be richer and have more meaning when we offer it to others.

3.1.7 - Three levels of self-reflection through voice

In these explorations we tried to always enrich the self-reflected image and we discovered our path through a flow that rises in its level of self-reflection.

The flow of our explorations started from a very physical, embodied and extroverted type of self-reflection, followed by a more social and networked type of self-reflection through our interaction with others when participating in the human social experience, finally to arrive at a more personal and introverted type of self-reflection.

In our work we don't see the voice as a simple sound, a simple process or a simple mechanism but as a full dimension or layer that has the potential to help people reflect on human existence. The voice as a full concept can be considered as this journey from the inner self to the outside world... and back.

3.2 - Voice minus Words

The first experimental project that explores the design space of the Vocal Self-Reflective system/methodology is VoiceminusWords (VmW). The project is a real-time signal processing software application that removes the verbal content from spoken sentences while keeping the individuality, expressivity and prosody of the message.

The goal is to offer an experience to shift people's attention from the words said to how the words are said. And by giving access to the speech normalized from the words and the literal content, this project enables people to focus on what remains.

When the text is made into gibberish, little bits of the discourse are actually shuffled in such a way that the audience cannot distinguish words but the qualities of the voice stay untouched: identity of the speaker, vocal posture, intonation, timbre, volume and emotional content).

This system was used in a performance piece presented as part of the *WOOD - WATER - WHISPER - WILD!* performance for the *Another Festival* in December 2013. The piece, Babel or not Babel is an oral recitation of the first 7 verses of Genesis. At specific moments in the reading, the performer interactively makes the meaning of the text disappear.

3.2.1 - Background

In this section, the background for and related works that inspired VoiceminusWords are presented. We present both inspirational artistic examples and technical methods to obtain the desired effect on the voice.

Maybe a main inspiration source for the VmW system and the Babel or not Babel piece comes from Charlie Chaplin's reaction to Talkies. In Modern Times (1936) Chaplin is obliged to stand and talk at two different symbolic levels: In the plot the main character has to perform a song but can't remember the lyrics and metaphorically as owner of a movie studio, Chaplin has to adapt to technological progress. The solution found by the filmmaker was to indeed use his voice but sing a gibberish song using a deliberately incomprehensible mixture of languages. The story told is made clear only by the acting and the tone of voice but no semantic meaning is present in the words themselves.

In The Great Dictator[119], Chaplin reiterates this process in the Dictator speech. Chaplin performs two famous speeches in this movie from 1940, a time of American neutrality. The same actor embodying two different

[119] Charles Chaplin and Paulette Goddard. *The Great Dictator: Film.* 1940

personas gives two entirely opposite discourses. One being the great dictator, the violent discourse of a conqueror said in another gibberish language and the second is a discourse in English for the freedom of all men. This extreme use of the voice reveals very efficiently its capacity to carry meaningful thoughts.

This example accentuates the importance of the voice in politics and public speeches, and the necessity of deciphering of this power of the voice. It was "the Führer's voice" (much more so than his face) that reached and thereby constituted "the German people," as one propaganda slogan put it[120].

Gibberish type languages have also been used to develop empathy between students[121], and has been shown to increases multicultural awareness. Gibberish is also commonly used in theater settings to train actors to develop more organic responses[122]. We touch here the search for something both personal and social in the voice, and removing the verbal content helps to access it. It can be in the prosody, in the vocal posture, in the breathiness, anything the performer chooses to express. In those performative situations, the voice may seem unnatural and lots of effort can be put in the choice of random sonority. Our system enables the performer to only focus on what he has to say and extract from that how he says it.

When it comes to the perception of Gibberish, the brain is very effective at detecting human voice among other sounds, and aside from the meaning of what is said, specific brain modules are associated with the recognition of speakers' identity and emotional states[123]. As already mentioned in the background section, the brain semantically analyses nonverbal inputs coming from the voice in the absence of understandable words[124].

The VOICEminusWORDS project takes its implementation inspiration from the methods used in data obfuscation designed for privacy. In terms of speech obfuscation, the system designed by Nicholas Joliat, Brian Mayton and Joseph A. Paradiso for anonymous audio browsing (Conference & Display, 2013) has been used as a starting point to design our voice-processing app. Because it required a real time reactive application, the system used in this thesis presents several differences from the original one, such as the output signal modulation by the input envelope to make the system seem more connected and reactive; or the choice of buffer reordering based on feature extraction and not on random computation.

[120] Victoria O'Donnell and Garth S Jowett. *Propaganda and persuasion*. Sage Publications, 1986; and Adrian Daub. " hannah, can you hear me?"—chaplin's great dictator," schtonk," and the vicissitudes of voice. *Criticism*, 51(3):451–482, 2009

[121] Ellen N Junn, K Morton, and Ivy Yee. The "gibberish" exercise: Facilitating empathetic multicultural awareness. *Journal of Instructional Psychology*, 22(4):324–329, 1995

[122] Viola Spolin. *Improvisation for the theater: A handbook of teaching and directing techniques*. Northwestern University Press, 1999

[123] DG Doehring and BN Bartholomeus. Laterality effects in voice recognition. *Neuropsychologia*, 9(4):425–430, 1971; and George Papcun, Jody Kreiman, and Anthony Davis. Long-term memory for unfamiliar voices. *The Journal of the Acoustical Society of America*, 85(2):913–925, 1989

[124] Katharina von Kriegstein, Evelyn Eger, Andreas Kleinschmidt, and Anne Lise Giraud. Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain research. Cognitive brain research*, 17(1):48–55, June 2003

3.2.2 – Interaction Design

The system was specifically designed to be used for the performance of *Babel or not Babel*. In this context, the performer recites a text and can at any time make the words become incomprehensible by pressing a pedal. While he/she continue to recite the text normally. This way, the piece performance alternates between times when the audience understands what is said, and times when the audience only hears how the text is said. In this case, the visual components (facial clues, body language, etc) also serve as communication clues.

In a slightly different context, the user can also choose to listen to recorded talks of speeches. By revisiting famous historic or political talks, one can experience the power present in the non-verbal elements still showing through the scrambled voice, once the verbal content is removed.

One other possible way to use this technology is to speak while listening to one's own voice scrambled in real time. The brain is used to attenuating the activity of the auditory cortex during phonation, in that we know what audible sound to expect[125]. Through the use of this technology, our own voice disrupts our cognitive expectation. This different way to experience speech returns the quality of your own voice to the foreground.

Upstream, the user can set two parameters depending on the quality of his voice and performance. Firstly the user can decide on the average size of shuffled buffer: between 100 and 500ms depending of the general tempo of the talk. Secondly the user sets a parameter for consonants detection between two modes: female voice or male voice.

3.2.3 - Implementation

The software was implemented using Max/MSP. It can be controlled either using the sliders in the window or using a midi device.

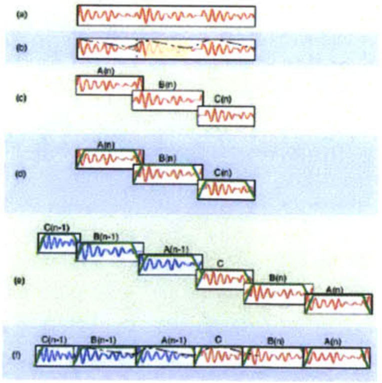


Figure 17: Stages of signal processing (a) original buffer (b) extraction of the signal envelope and if possible detect vowels (in beige) and onset of consonants (in dotted line) (c) split between buffers A, B and C guided by vowel and consonants detection, non random, around a chosen buffer size (d) windowing (e) shuffle buffers A and B after the real time buffer C (f) multiplication by the original amplitude envelope

Figure 17 shows the different stages of signal processing. The first step of the signal processing is to chop the signal into 3 small buffers A,B,C and to play them back in the order C,B,A with C played in real time. While C is being played, the recorded buffers are multiplied by an envelope signal to enable a smooth reconstruction without clicking sounds. By trial and error, we determined that the optimum size to keep the voice quality while keeping the words incomprehensible is between 100ms and 500ms.

If the buffer size is always constant, the result is not convincing for two reasons. First the constant time between buffers changes creates a "beat effect". Indeed, the sound output oscillates at a constant frequency corre-

---

[125] John F Houde, Srikantan S Nagarajan, Kensuke Sekihara, and Michael M Merzenich. Modulation of the auditory cortex during speech: an meg study. *Journal of cognitive neuroscience*, 14(8):1125–1138, 2002

sponding to the buffer size. To solve this problem, we add some randomness in the buffer size to have a slightly different buffer length at each ABC-cycle.

The second issue comes from the disconnection between the facial cues (breathing and lips motion) and the audio. For the sound and visual to not feel desynchronized, the visual clue of a pulsation has to be mapped to a pulsation while sustained sounds such as vowels or silences should be respectively mapped to vowels and silences. Because of this necessary mapping, the system needs to be aware of the difference between silence, vowels and consonants (especially consonants of the stops type /p, t, k, b, d, g/)[126]. To deal with the silences, we added a threshold in the input that deactivates the shuffling of buffers when the signal is too weak. From the vocal input in real time we also detect consonants types using the zsa.descriptor package for speech description[127] and we split the buffers immediately before consonants sounds when possible. Indeed, we want the lips' motion to roughly correspond to the sound variation. To obtain this, the system detects the rising edge of consonant onsets, and plays a different buffer starting also with a consonant sound.

In order to smooth the irregularities, the resulting signal is multiplied by the envelope of the real time input. This also contributes to obtaining an output even more strongly correlated with the performance.

3.2.4 – Execution and result

This VmW system was used in a performance piece presented as part of the *WOOD - WATER - WHISPER - WILD!* performance for the Media Lab arts and design festival in December 2013. The piece calles *Babel or not Babel* is an oral recitation of the 7 first verses of Genesis. At specific moments in the reading, the performer interactively activates the WmV system to make the meaning of the text disappear.

Aside from a wrong start at the beginning of the performance due to forgetting to unmute the system, the software worked as expected during the 6 minutes of the piece.

From the reaction of the audience members to the performance, it appeared that the first contact people have with this technology results in surprise and disorientation. People reported that it generally took them a couple of seconds to understand that the semantic meaning is scrambled. People also reported they felt disconcerted by the strange connection between the natural and constant attitude of the performer and the oddness of their sonic experience. One person told us she felt "like in the type of dream when people around you speak a language you don't know but you

[126] Pierre C Delattre, Alvin M Liberman, and Franklin S Cooper. Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, 27(4): 769–773, 2005

[127] LR Rabiner and RW Schafer. Digital speech processing. *The Froehlich/Kent Encyclopedia of Telecommunications*, 6:237–258, 2011; and Mikhail Malt and Emmanuel Jourdan. Zsa. descriptors: a library for real-time descriptors analysis. *Sound and Music Computing, Berlin, Germany*, 2008

still understand what is going on, or when you read a text in your dream where the letters are all mixed up but you can still read it".

This first instance of the Self-Reflective Vocal Experience plays on the social context of the voice. The mirror that is offered here is a normalization of the voice from the verbal content. By giving access to the speech normalized from the words and the literal content, this project enables people to focus on what remains. Either as an interactive game or as an audience member watching a performance, this system takes any familiar voice and transforms it into something both familiar and strange, both distant and close. This project is a first possible vocal mirror in the way that it split the voice into two elements, the message and what conveys it. The only part that is reflected to the user is the conveyor without the message. The message is absorbed and disappears.

3.3 - V^3 - Vocal Vibrations Visualization

V^3 is a system that offers a real-time visualization of vocal sound vibration pattern. For this project, we developed the hexauscultation mask. Hexauscultation is a portmanteau word between the Greek word for the number six "hexa" and "auscultation"the action of listening to the internal sounds of the body: a head set sensor that measures the bioacoustic signals from the voice at 6 vocalized points in the face and neck. Those measured signals are sent and processed by our software to offer a visual representation of the vibration intensities at the 6 points of the face as shown in figure 18.

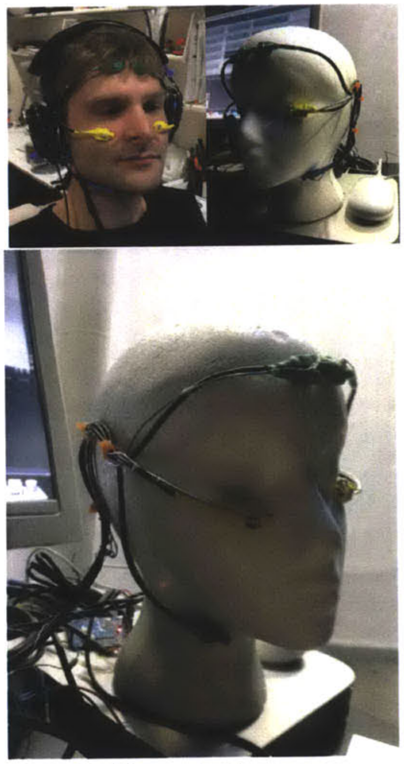


Figure 18: Hexauscultation mask

Auscultation is the act of listening to internal sounds from the body. The sensor built for this project listens at the skin surface to the internal human voice from 6 different locations. From this listening, our system establishes a direct and instinctive link between the physical source of the sound and the sound itself. By trying different sound textures (vowels, humming, consonants, etc) the user has a direct visual feedback of where the sound texture comes from.

According to the source-filter model of vocal production, the filter part of the vocal apparatus, called the vocal tract shape, defines parts of the sound quality. The differences between two sounds,"a" and"u"for example, are controlled by muscle contraction states, opening and closing of cavities and placement of the tongue. All these elements are shaping the vocal tract and create a unique resonant characteristic that results in a unique formant pattern, which is to say, in a unique sound. The V^3 project gives access to some upstream physical control parameters of the voice by making the sound origin more accessible.

This project decomposes the "filter" part of the vocal sounds we produced and makes it visible. This point of view offers an alternative consideration of a potential "fullness" of the voice that is not guided by phonetic rules or the grammar and sound associated with a specific language but guided by the physical control offered our body.

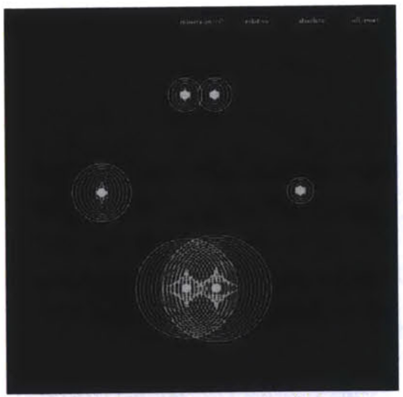


Figure 19: Interface of the V^3 software without vibration input in relative mode

3.3.1 - Background

The idea of considering the voice from the point of view of vibrations and to make those visually accessible was guided by different prior work sources. We already presented in the second chapter the field of bioacoustics and auscultation. The V^3 projects draws its roots from those ideas. In this section, we will present an insight into two other inspirational domains: voice visualization methods and voice related headset devices.

From prior work in voice visualization, we already mentioned in the Background section some art work form Golan Levin. His work tackles the visualization of voice in several artistic projects in many different axes. Some projects such as Re:Mark or Ursonate use letters of the alphabet to represent sounds changing shape and position on the screen. In the messa di voce project, visuals are more abstract and engage the public into thinking about their voice as an instrument rather than in words. The visuals are engaging and different versions are developed both for amateurs and professional virtuosic vocalists[128]. The parameters controlling the visualization (pitch, spectral content, and autocorrelation data) are tightly linked with what we perceive as the sound quality. The great organic feel of Levin's work comes from the use of physical modeling techniques in the visualization including particle systems, elastic spring meshes, fluid simulations or cloud simulations. In our work, though the design aspect is less present, the liveness is not added afterword in the process, rather, we intend to not break the thread of organicity coming from the voice. In the context of speech visualization, the examples of conversation clock[129] and sonic shapes[130] offer a comprehensible view of human vocal exchange in terms of sonic parameters with the goal of visualizing communication patterns between people in different kinds of relationships and cultures. Cematics are an example of visualization of the physicality and sonic vibrations. Through the cymascope projects, they have been used to visualize vibration patterns from the human voice. Figure 20 shows the cymatic visualization of vowels pronounced by Vera Gadman[131].

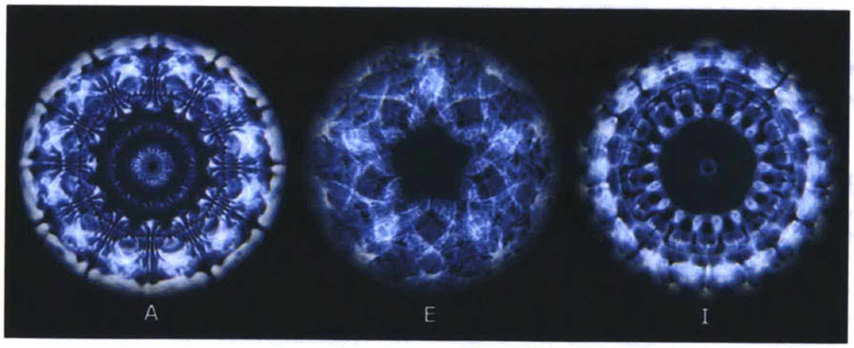


Figure 20: Cymascope of the sound A from a female voice by Vera Gadman

On the design and characteristics of the sensors used in bioacoustics, the importance of limiting the air-borne sensitivity of the sensors to only measure the tissue-borne signal pledge in favour of piezzo-electric sensor instead of other accelerometer-type technologies or microphones[132].

The idea of a mask designed to access the voice and enhance or transform it has been used in different domains. Wearable systems used for

[128] Golan Levin and Zachary Lieberman. In-situ speech visualization in real-time interactive installation and performance. In *NPAR*, volume 4, pages 7–14, 2004

[129] Judith Donath, Karrie Karahalios, and Fernanda Viegas. Visualizing conversation. *Journal of Computer-Mediated Communication*, 4(4):0–0, 1999

[130] Mary Pietrowicz and Karrie Karahalios. Sonic Shapes: Visualizing vocal expression. *ICAD*, pages 157–164, 2013

[131] Cymascope. Cymascope, sound made visible. URL `http://cymascope.com/cyma_research/phonology.html`

[132] Matías Zanartu, Julio C Ho, Steve S Kraman, Hans Pasterkamp, Jessica E Huber, and George R Wodicka. Air-borne and tissue-borne sensitivities of bioacoustic sensors used on the skin surface. *Biomedical Engineering, IEEE Transactions on*, 56(2): 443–451, 2009

voice transformation have been used in many cultures for ritual or artistic purposes. The mask that actors wore during theatrical performances in ancient Rome was called the Persona ("speaking through"). It was both a way to display their role and an acoustic tool to carry their voice in the theater[133]. Other example are the alaspraka masks of Papua New Guinea and the African mirliton[134]. Those systems use the shape and material of the device to filter the voice in a way quite similar to the vocal tract in the common source-filter model[135]. Nowadays wearable technological devices are used for vocal performance enhancement and transformation[136] The transformation or an actor persona speaks through his voice, through technology.

3.3.2 – *Interaction Model*

V^3 has a simple interaction model: the user puts the two sides of the mask on his face supported by the ears. If necessary, due to the shape of the face, small hypoallergenic double-sided tape can be used to keep the sensors in contact with he skin. The user then puts on a microphone and headphones. Every time the user produces a sound, the system visualizes in real time the amplitude of the vibrations at the six locations. Those are displayed on the screen under she shape of concentric waves whose amplitude increases. The position of the vibration points on the display window can be modified by hand to match better their position on the person's face.

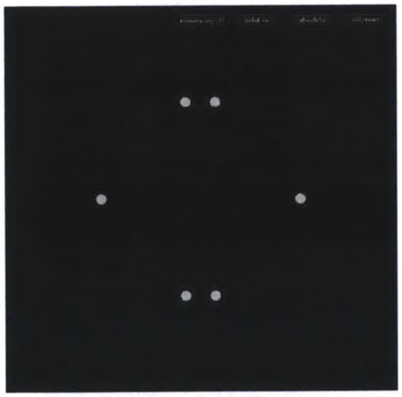


Figure 21: Interface of the V^3 software without vibration input, buttons camera on/off; absolute mode; relative mode

We call vibration patterns the 6*1 vector formed by the signals. Because the variation with volume appeared to be relatively linear, we can normalize the vector to become independent from the general volume of the sound to focus on the variation between the points. The system contains 2 viewing modes: absolute and relative. In the absolute mode, the size of the ring corresponds directly to the intensity at that point. In the relative mode, the vibration pattern is normalized by the mean of all the points. This mode is sometimes more relevant to understand the correlation and variation of the pattern and to link the pattern to differences in sound production. When using a webcam, the system also offers a "mirror" mode in which the vibration rings are displayed on top of a real time video feed.

3.3.3 - *Implementation*

Hardware

The sensors are piezzo based phantom powered vibration sensors. The sensors are adapted from Bryan Mayton's design for underwater microphone from the Tidmarch project[137]. For this project, we used small 1/4

[133] R. D. Cherry. Ethos Versus Persona: Self-Representation in Written Discourse. *Written Communication*, 5(3):251–276, July 1988

[134] Sidney Littlefield Kasfir. *West African masks and cultural systems*, volume 126. Musée royal de l'Afrique centrale, 1988; and Philip M Peek. the sounds of silence: cross-world communication and the auditory arts in african societies. *American Ethnologist*, 21(3):474–494, 1994

[135] BM Blackwood and Henry Balfour. Ritual and secular uses of vibrating membranes as voice-disguisers. *Journal of the Anthropological Institute of Great Britain and Ireland*, pages 45–69, 1948

[136] Alex Stahl and Patricia Clemens. Auditory Masquing : Wearable Sound Systems for Diegetic Character Voices. (Nime):15–18, 2010; and Mel Goldberg. Face mask with voice modifying capability, July 28 1987. US Patent 4,683,588

[137] Media Lab Responsive Environments Group. Tidmarsh living observatory. URL `http://tidmarsh.media.mit.edu/`

inch piezzo sensors whose single is preprocessed by a phantom power preamplifier board and the output is an audio signal read by a 6 input channel audio interface (we us a TASCAM US-1800). The sensors are designed to be very small and light. We printed and mounted the components at the lab, chassis grounded the board on the metal thread, then isolated and shielded the system. For questions of safety, comfort, convenience and aesthetics, we wrap each sensor into shrink-wrap. This ensures that no shorting can happen by handling the boards, secures the link with the XLR cables, and provides a comfortable surface for contact with the skin.

The sensors are mounted on flexible metal wires that are shaped to support the sensors at the right place, in contact with the skin and supported behind the ears. We tried different ways to ensure good contact between the skin and the sensor. We used very thin double-sided tape to reduce the damping of the vibration and to ensure very good contact. The tape used is hypoallergenic specially made for skin. The metal structure is made by bending an aluminum wire to the correct curve to fit the face.

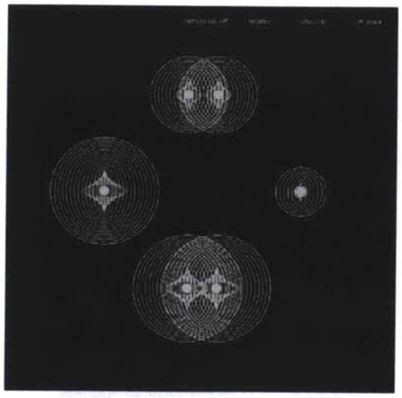


Figure 22: Example of vibration patern observed with the V^3 system

Software

The real time audio processing part of the system is done using Max/MSP. Our patch pre-filters the 6 input channels and extracts their envelopes. Then we assign a different coefficient on the three levels of measurement: throat, cheek and forehead. Finally, we send the 6 envelopes in real time by Open Sound Control (OSC) to the visualization part of the system.

The visual parts of the software are made in Java. It receives the intensity of the signal in real time and displays it. The location of the 6 points are by default displayed in pseudo circle but can be modified manually.

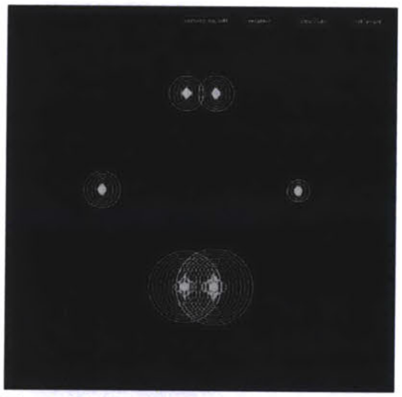


Figure 23: Example of vibration patern observed with the V^3 system

3.3.4 - Evaluation

The first version of the structure didn't use metal but only tape to keep the sensors in place. This solution was not satisfactory because the sensor units were pulled out from the face by their own weight and the tension of their cables. The addition of the metal structure solved this problem. The next challenge was to find a mask shape that could fit any facial physiology. This was done by trial and error and by opting for a thick but very flexible metal that can be slightly reshaped for each user. The system proved quite good at not picking up the facial motion when the user moves his whole face or when articulating sounds. Individual differences in skin thickness and dryness, hairiness and ear position seemed to cause slight differences in the signal sensed by the sensors while remaining in a reasonable range.

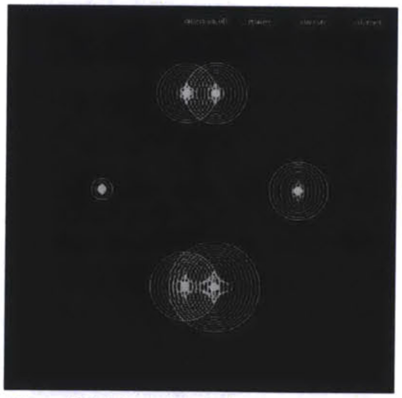


Figure 24: Example of vibration patern observed with the V^3 system

Many people have been involved in testing the system asked to vocalize freely and to give informal feedback. Comments about the structure reveal that people find the measurement system very light and not annoying. Even though it is sometimes fastidious to put on the people, the participants reported they forgot about it after a couple of minutes. Some people reported being frustrated that they were not able to make the top of their head vibrate. A professional singer reported really liking the device and thinking it could be helpful in vocal training sessions. During the prototyping process, while playing back an audio track recorded through the sensor device, a colleague reported that though the voice was perfectly recognizable as being the author's, there was something different in it, as if it sounded like that voice had something strangely familiar, like the way she was hearing her own voice in her own head.

3.4 - The ORB in Vocal Vibrations

3.4.1- Introduction

The Vocal Vibrations installation (see figures 26 and 30) is an interactive voice-based multisensory musical installation. It is the result of a collaboration between The Opera of the Future group at the MIT Media Lab, the Dalai Lama Center for Ethics and Transformative Values at MIT and Le Laboratoire, an art and design center based in Paris[138].

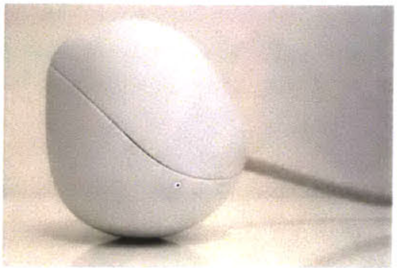


Figure 25: The ORB (photo by Raphael Cei)

The experience offered through the installation seeks to engage the public in thoughtful singing and vocalizing, while exploring the relationship between human physiology and the resonant vibrations of the voice. This installation consists of a space for reflective listening to a vocal composition (the Chapel) and an interactive space for personal vocal exploration (the Cocoon). In the interactive experience, the participant also experiences a tangible exteriorization of his voice by holding the ORB, a handheld device that translates his voice and singing into tactile vibrations. This installation encourages visitors to explore the physicality and expressivity of their voices in a rich musical surrounding. In the context of the installation, the ORB helps people gain awareness of the physicality of their voice through vibrations. Figure 26 shows the ORB project in the context of the Vocal Vibrations installation

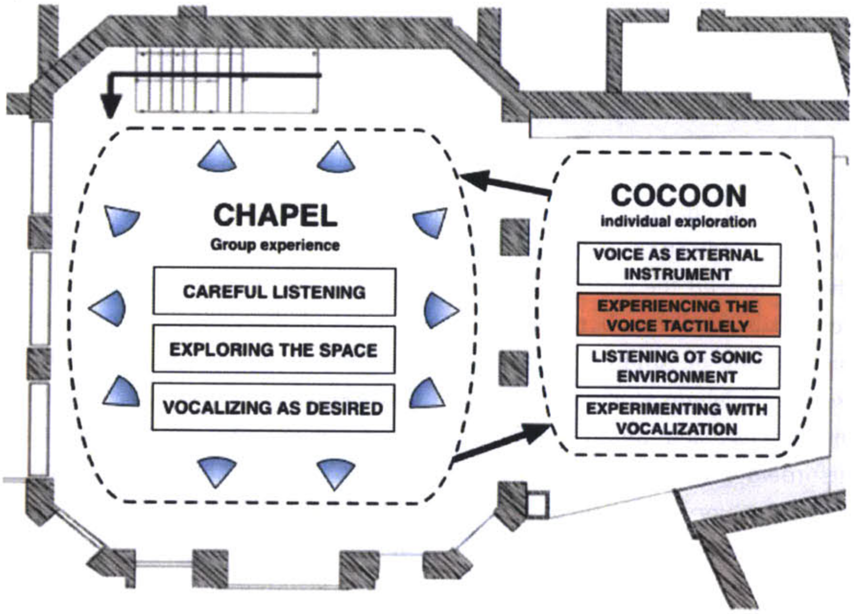


Figure 26: The ORB project in the context of the Vocal Vibrations instalaltion

---

[138] La laboratoire. URL `http://www.lelaboratoire.org/en/archives-18.php`

This project is the result of a collaborative work; we include it in this thesis because the initiative aims to offer to the public a new experience of their voice. Although this section covers the overall production of the project, the primary focus is on the author's involvement and work carried out in the context of the installation.

Details on this project can be found in the publication that resulted from it NIME 2014 or on the project webpage[139]. In the following paragraph, we go back over the user interaction, the description of the different implementation stages, the results and some elements of evaluation.

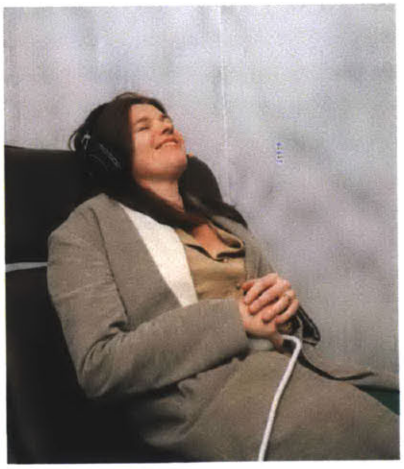


Figure 27: Participant in the cocoon interacting with the ORB (photo by Tod Machover)

3.4.2 - Background

We have already covered some of the related works about voice-based installation and research. In this section we reiterate some of the inspiring literature for the overall initiative and the orb specifically.

The initiative

The act of singing and vocalizing creates vibrations throughout the body. However, people are generally not aware of or focused on these vibrations. The awareness of vocally-produced vibrations can be a source of meditative focus, as well as a way for everyone from novices to trained singers to understand their instrument better. Vocal coaches, singers, and voice professionals have very rich terminology to characterize different voices. However, because the vocal apparatus is hidden from sight, the vocabulary that is used is abstract and hard to grasp for the non-initiated[140]. The focus provided by tactile and physical feedback can help to give intimate, while still objective, access to the voice. Figure 30 gives a panoramic view of the overall installation in Paris.

One of the main motivations for this project comes from the ancestral tradition of chanting and mantra and the role played by voice in many forms of meditation and health related practices[141]. Section 2.2.3 of this thesis covered the wellbeing benefits that the voice can bring. However, very little work has been done on the effects of the vibrations produced in the body by singing, or on the relaxation and meditation potential of the voice. Many studies have shown that meditation training (especially mindfulness meditation) may be an effective component in treating various disorders such as stress, anxiety, and chronic pain[142]. Despite the voice being a major part of several meditation traditions, the effects of the voice in meditation are mostly unexplored. In one study, medical imaging has shown that cerebral blood flow changes during meditation that incorporates chanting on resonant tones, in ways that cannot be explained solely by breath control[143].

---

[139] Vocal vibrations. URL `http://opera.media.mit.edu/projects/vv/`

[140] Marianne Latinus and Pascal Belin. Human voice perception. *Current biology : CB*, 21(4):R143–5, February 2011

[141] David B Wolf and Neil Abell. Examining the effects of meditation techniques on psychosocial functioning. *Research on Social Work Practice*, 13(1):27–42, 2003; and Sanjay Kumar, HR Nagendra, NK Manjunath, KV Naveen, and Shirley Telles. Meditation on om: Relevance from ancient texts and contemporary science. *International journal of yoga*, 3(1):2, 2010

[142] Richard J Davidson, Jon Kabat-Zinn, Jessica Schumacher, Melissa Rosenkranz, Daniel Muller, Saki F Santorelli, Ferris Urbanowski, Anne Harrington, Katherine Bonus, and John F Sheridan. Alterations in brain and immune function produced by mindfulness meditation. *Psychosomatic medicine*, 65(4):564–570, 2003; and Jon Kabat-Zinn, Leslie Lipworth, and Robert Burney. The clinical use of mindfulness meditation for the self-regulation of chronic pain. *Journal of behavioral medicine*, 8(2): 163–190, 1985

[143] Bangalore G Kalyani, Ganesan Venkatasubramanian, Rashmi Arasappa, Naren P Rao, Sunil V Kalmady, Rishikesh V Behere, Hariprasad Rao, Mandapati K Vasudev, and Bangalore N Gangadhar. Neurohemodynamic correlates of 'om'chanting: a pilot functional magnetic resonance imaging study. *International journal of yoga*, 4 (1):3, 2011

Figure 28: Vocal Vibrations instalaltion (photo by Tod Machover)

Vibration therapies have shown many different medical effects depending to its frequency, mode of exposure, length of application, amplitude and acceleration. Several of those methods use frequencies that are in the human vocal range (80 to 1200Hz). For instance, vibrations of 50 to 150Hz have beneficial effects on chronic pain[144]. 100Hz vibrations have been shown to reduce tooth pain [145]. In the biological scale, vibrations around 300Hz are used to enhance DNA synthesis[146]. Lower frequencies from 10 to 45Hz have positive effects on fluid [147] and blood [148] circulation and thus help reduce the risk of edema formation[149]. Vibrations around 15-30Hz have also been used to help people with Parkinson motor impairment.

Tactile Hearing

The development of the concept and design of the ORB was inspired by previous work in devices designed to exteriorize inner physical and mental states. The fields of experience design and pervasive technology offer examples of devices that exteriorise aspects of the individual that happen below the conscious threshold[150]. By exposing and revealing those aspects, they can help the user engage in a reflective process. Designed by Kelly Dobson, Omo acts as a relational machine[151] by imitating the breathing rhythm of the user. The pervasive mirror is a mirror-like display that reflects an image[152] of what the user will physically look like, in several years, given their current aspect and health-related habits.

In the case of the ORB the object remains visually unchanged and does not produce sound; it is designed to speak to the skin. The skin is the largest organ of the body, about 2,500 cm2 in the newborn and about 18,000 cm2 for adults. It represents 18 percent of our total body weight[153]. The choice of connecting vocal sounds and the sensations of the skin was also inspired by previous research on the Tactaid, a tactile

---

[144] TC Lundeberg. Vibratory stimulation for the alleviation of chronic pain. *Acta physiologica Scandinavica. Supplementum*, 523:1–51, 1982; and Régis Guieu et al. Analgesic effects of vibration and transcutaneous electrical nerve stimulation applied separately and simultaneously to patients with chronic pain. *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques*, 18(2):113–119, 1991

[145] P Hansson et al. Influence of naloxone on relief of acute oro-facial pain by transcutaneous electrical nerve stimulation (tens) or vibration. *Pain*, 24(3):323–329, 1986; and Anders Ekblom and Per Hansson. Extrasegmental transcutaneous electrical nerve stimulation and mechanical vibratory stimulation as compared to placebo for the relief of acute oro-facial pain. *Pain*, 23(3):223–229, 1985

[146] Jie Liu et al. Biosynthetic response of cultured articular chondrocytes to mechanical vibration. *Research in experimental medicine*, 200(3):183–193, 2001

[147] Julian M Stewart et al. Plantar vibration improves leg fluid flow in perimenopausal women. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 288(3):R623–R629, 2005

[150] Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):5, 2002

[151] Kelly Elizabeth Dobson. Objects that interact with a user at a visceral level, October 1 2013. US Patent 8,545,228

[152] Ana C Andrés del Valle and Agata Opalach. The persuasive mirror. In *Proc. Int'l Conf. Persuasive Technology*, 2006

[153] Ashley Montagu. Touching: The human significance of the skin. 1971

hearing aid that codes sound information via a set of vibrators resting on the forearm. Studies have shown that those types of coded vibrations enhance lip-reading performance in hearing impaired individuals[154].

In terms of previous work on how to give the skin access to tactile vibrating sensations, different methods have been used. The use of small motors is common to create vibrations, as in different types of cellphones. If the amplitude of the vibrations can be determined quite precisely, their frequency range is generally too limited to provide convincing expressivity qualities[155]. Small audio transducers have been used before to transmit vibration information such as in the ComTouch device, a remote communication device that augments the voice with vibrations depending on the hand pressure on the object[156]. Reverse electrovibrations consist of sending a very weak electric signal to a person to create a tactile sensation via the specific field created around his or her skin. This effect has been used to program virtual texture on real objects[157]. The object needs to contain an electrode and an insulator.

3.4.3 – User interaction / User experience

The Vocal Vibrations installation consists of two connected spaces that encourage participants to experience and explore the singing voice, especially their own voices, in thoughtful ways.

Figure 29: The chapel, user interaction (photo by Julien Benayoun)

The Chapel

When participants first arrive at the Vocal Vibrations installation, they enter a communal space designed for close listening. In this space, which we call the Chapel, the audio is a pre-composed electroacoustic composition by Tod Machover based on many layers of solo and choral vocal material,

---

[154] Karyn L Galvin et al. A comparison of tactaid ii and tactaid 7 use by adults with a profound hearing impairment. *Ear and hearing*, 20(6):471, 1999

[155] Angela Chang, Sile O'Modhrain, Rob Jacob, Eric Gunther, and Hiroshi Ishii. Comtouch: design of a vibrotactile communication device. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 312–320. ACM, 2002

[156] Angela Chang, Sile O'Modhrain, Rob Jacob, Eric Gunther, and Hiroshi Ishii. Comtouch: design of a vibrotactile communication device. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 312–320. ACM, 2002

[157] Olivier Bau, Ivan Poupyrev, Mathieu Le Goc, Laureline Galliot, and Matthew Glisson. Revel: programming the sense of touch. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 2785–2786. ACM, 2013

designed such that a D is almost always present in the score. The musical material used in the composition comes from a broad spectrum of vocal traditions including Tuvan throat singing, early Renaissance choral music, and solo soprano explorations. At any time, participants can vocally follow along with this composition by singing a D. Headphones playing a pure D note, sung at different octaves, are also available in the space to help the participant to find a note that fits comfortably in his or her range. The installation is made up of 10 high-fidelity Bowers Wilkins speakers on tripods that have been configured in a large oval within the room. Each speaker plays different elements of the composition providing a dynamic surround experience.

The Cocoon

In the second portion of the installation, a private environment, the Cocoon, allows individual visitors to explore new vocal sounds and the vibrations generated by their own voice. From the Chapel, an assistant guides each participant into the interactive experience. A short "training" session follows, in which the participant is encouraged to take the first step into producing vocal sounds. The participant is asked to hold the D and is given simple guidance to explore a range of vocal variations on a single pitch, such as different vowels, sounds, rhythms, textures, and timbres. We seek to free participants to experiment with a wide range of sounds. We designed the system such that there is no "right" or "wrong" way to use it. Whatever sound a participant produces does not degrade the musicality of the experience. People are invited to vocally follow the music they hear, but they barely hear their own voice through the headphones, which enables them to feel more comfortable trying new sounds.

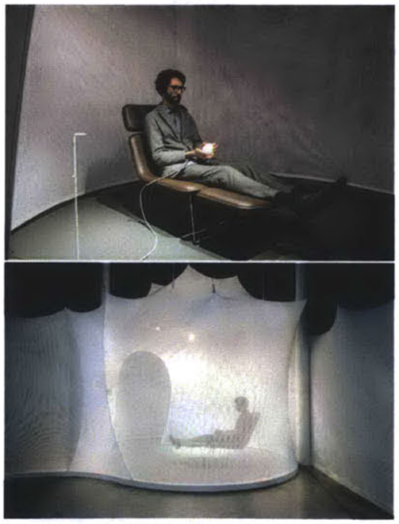


Figure 30: The cocoon view from the chapel (photo by Julien Benayoun)

To provide awareness of physical processes involved in the vocal production process, the ORB maps the voice signal into tactile vibrating sensations. Because fingertips contain more sensory receptors than our vocal vibrating chamber[158], holding the ORB while vocalizing gives access to the voice through the physical effects it causes in our body

The ORB

By exploring different types of vocal sounds, the user feels differences in the dynamic, location and intensity of the vibrations.

Voice can be seen as composed of prosody and timbre (source/filter). The timbre and its variations are the consequence of the shape of the vocal apparatus: opening and closing of the vocal and nasal cavities and the position for the tongue[159]. As mentioned earlier, this shape makes the head a specific filter by mechanically conducting the wave sound to different

[158] Tatsuya Kitamura. Measurement of vibration velocity pattern of facial surface during phonation using scanning vibrometer. *Acoustical Science and Technology*, 33(2):126–128, 2012; and S Levänen and D Hamdorf. Feeling vibrations: enhanced tactile sensitivity in congenitally deaf humans. *Neuroscience letters*, 301(1):75–7, March 2001

[159] Kenneth N Stevens. *Acoustic phonetics*, volume 30. MIT press, 2000

places and having the head resonate more at different frequencies. It is the shape of the conduit and so where the resonant vibrations take place that determines the frequency spectrum of the emitted sound.

The vibrational behavior of a solid sphere was first studied by Siméon Denis Poisson in 1829[160]. Two centuries later, physicists can simulate the precise vibrational behavior of thick hollow spherical objects through physical modeling[161]. For the development of the ORB, the material, shape and thickness were not established from precise computation but as a combination of scientific, ergonomic and aesthetic considerations.

[160] Siméon-Denis Poisson. *Mémoire sur l'équilibre et mouvement des corps élastiques.* L'Académie des sciences, 1828

[161] Jae-Hoon Kang and Arthur W Leissa. Three-dimensional vibrations of thick spherical shell segments with variable thickness. *International journal of Solids and Structures*, 37(35):4811–4823, 2000

Figure 31: The ORB (photo by Julien Benayoun)

When feeding back this sound in a non-perfectly round hollow object, the same phenomenon happened in reverse. The frequencies excite the material differently at different location dues to the curvature of the shape and the possibilities for the sound waves to couple with the object shape. By feeding different sound signals in the ORB, we make the object resonate the most at different locations that are linked to the location happening on the head of the person vocalizing.

Through this object, the installation offers users a tool to exteriorize their voice and experience another form of connection with it, as well as to engage with their voice as one engages with an external instrument.

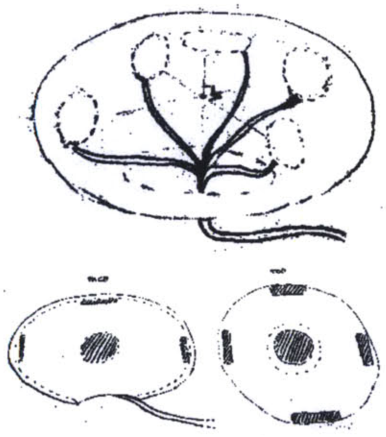


Figure 32: Sketches for the first prototype of the ORB

3.4.4 – *Implementation*

The constraints in designing the ORB were to build an object to exteriorize the voice while not losing the feeling of connection. Though the choice of building a vibrating device was upstream in the process the final characteristics and behavior of the object required several prototypes both in the hardware and software sides.

Hardware

On the hardware side, we explored several prototypes of sizes, materials and thickness for the shell. The walls needed to be thin enough to transmit vibrations and not damp too much particular frequencies. As the material needed to conduct the mechanical sensation without producing sounds, we eliminated wood, metal and plastic. The first prototype of the ORB uses a shell from a product manufactured by the Ambient Device Company[162]. This first shell was an ovaloid shaped frosted glass shell measuring about 15 by 15 by 10 centimeters.

[162] Pritesh Gandhi. Ambient device company. URL `http://www.ambientdevices.com/`

Figure 33: Pile of empty ORB shells

The final shape of the ORB is the result of a collaboration work with the Bold Design company located in Paris. The device is an almost ovoid porcelain shell measuring about 10 by 9 by 9 centimeters, with five transducers attached on the inside wall. The materials and precise settings are chosen to maximize the tactile feeling of vibration while minimizing any audible resonance from the device. The object can be held in the hands in any orientation and angle, with different positions varying the perceived vibrational effects on the hands and fingers. Because ceramic-type materials have the microstructural property of presenting no directional atomic order[163], the material offers the beneficial properties of smoothly and blending the vibration from one transducer to another while keeping certain localized effects. Those localized effects enable the signal to naturally vibrate more at certain points on the object given the frequency spectrum of the signal. The vibrating elements we choose are small 8 ohms audio transducers. Five of them are localized and glued inside the shell. The signal is sent from the computer to amplifiers that boost the signal and increase the amplitude of the vibrations.

Software

On the software side, our goal was to obtain rich vibration sensation and a feeling of connection and directness for every type of voice. This system designed to control the ORB behavior is a Max/MSP patch that sends a processed signal to each of the 5 localized channels based on a set of control parameters.

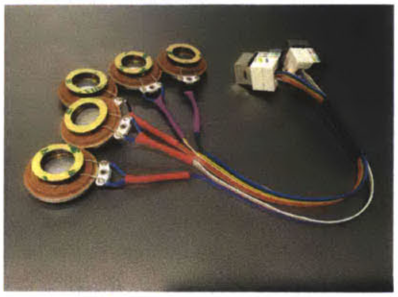


Figure 34: Transducer snake before being glued in the ORB shell

The system is connected to the microphone. We wanted the raw voice to be constantly present in the ORB. For this, one of the 5 transducers always contains the vocal signal coming from the microphone. Feeding the raw voice directly to the ORB transducer had several drawbacks that we fixed by preprocessing the vocal signal. First the signal contains lots of high frequencies that produce sounds in the shell while not producing perceptible tactile sensations. Those high frequencies also brought unpleasant feedback sound when the microphone is placed too close to the ORB. To fix this issue, we filter the input signal to remove high frequencies above 1 kHz to avoid feedbacking and also because fingertips cannot perceive those frequencies. Secondly, the human vocal range and especially the female human range mainly contain frequencies too high for the fingers to perceive. Women with high voices could only feel very weak vibrations from the ORB. To solve this problem, we create a second version of the real time signal that is pitch shifted an octave lower than the original. We then combine the original and the pitch shifter signals with a ratio 40% / 60%. For very low voices, this pitch shift doesn't change the tactile perception. For high voices, this part of the signal helps increase the perceived vibrations while keeping the signal very connected to the voice. This way, we make the vocal vibrations really tangible in the ORB independently of the characteristics of the participant's voice.

---

[163] William A Curtin. Theory of mechanical properties of ceramic-matrix composites. *Journal of the American Ceramic Society*, 74(11):2837–2845, 1991

Figure 35: User interface of the ORB behavior software

In addition to the raw vocal signal that is sent to the top channel for intuitive tactile feedback, the system allows the ORB to also vibrate with additional textures. Those textures can come from prerecorded audio samples, audio textures synthesized in the system by granular or additive synthesis, or from buffers corresponding to the vocal input of the user himself. The textures can be sent with specific location, speed and scattering around the surface, creating abstract impressions like purring, hatching, whirling, etc. The parameters controlling the textures can be predefined or be controlled in real time by external systems. The system can also add additional effects on the signal such as delay, attenuation and a light feedback. The mapping of the ORB was made to be very modular and extensible.

Figure 36: Behind the curtain, the ORB setting in the Vocal VIbrations installation

The system is extensible in a way that it can also extract parameters from the voice in real time. As the participant vocalizes in a microphone, the raw signal is analyzed and converted into a first level of control parameters: pitch, loudness, linearly averaged frequency spectrum, and harmonicity. These are computed by spectral analysis. The objective in this choice of parameters was to underline the feeling of an instinctive and immediate connection from the user to the system. These elements of the voice are perceived very strongly, so they can aid in creating an obvious link between vocal variation and the resulting output of a system. This section has been implemented with the Max/MSP software and uses custom made analysis tools as well as plugins from different sources: the grainstretch patch by Timo Rozendal, zsa.descriptors by Mikhail Malt and Emmanuel Jourdan[164] and the analyzer patch by Tristan Jehan[165]. The system can also be extended through external software communicating in real time via open sound control to map the vocal parameters into higher level parameters controlling the tactile effects in the ORB.

From this system many different parameters and behaviors were tested to obtain the final mapping used in the Paris version of the Vocal Vibrations installation. For the installation in Paris, we used part of those possibilities. Besides the raw voice in the top transducer, an additional effect consisted of sending the voice signal also to one of the 4 lateral transducers. While the participant vocalizes, the location of this additional channel changes. To control this effect, the low level vocal parameters are processed by an intermediary system called Expressive Performance Extension System that uses machine learning to decide when onsets from the voice would trigger change in the location of the second channel[166]. This program was developped as part of Elly Jessop PhD dissertation.

[164] Mikhail Malt and Emmanuel Jourdan. Zsa. descriptors: a library for real-time descriptors analysis. *Sound and Music Computing, Berlin, Germany*, 2008

[165] Tristan Jehan and Bernd Schoner. An audio-driven perceptually meaningful timbre synthesizer. *Analysis*, 2(3):4, 2002

[166] Elena Jessop. *The body parametric:Abstraction of Vocal and Physical Expression in Performance*. PhD thesis, Massachusetts Institute of Technology, 2014

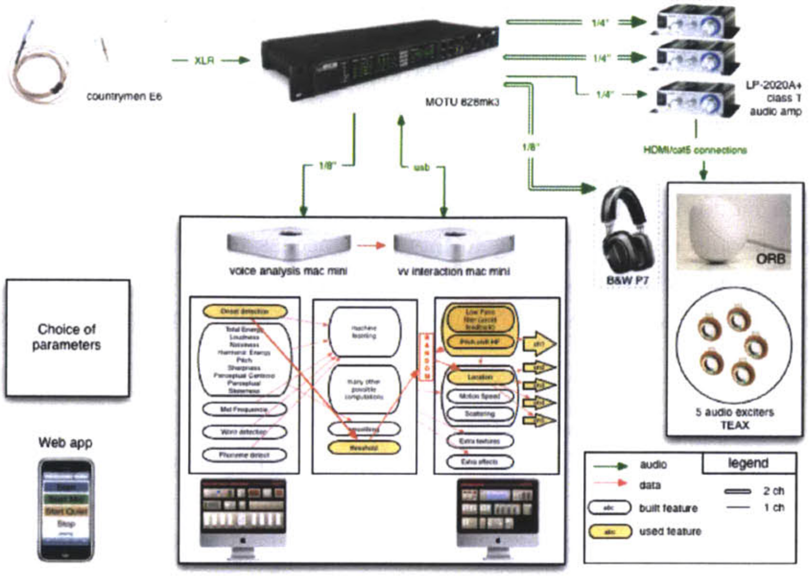


Figure 37: System Diagram of the ORB system as set up in the Vocal Vibrations installation

3.4.5 - Execution and results

Challenges

The set up of the installation took a week during which many different groups worked on different tasks such as the stenographic design, lighting, the installation of audio speakers, the acoustic tuning, the sound mixing. One expected challenge of this installation was to tune the audio levels and thresholds of the chapel and the cocoon to limit sound leakage. Indeed, the two sound spaces had to cohabitate. In order to not disturb the cocoon experiment, we had to choose the chapel level such that the cocoon participant does not hear the sound from the other room and also such that the microphone would not pick up and send to the ORB sound coming from the chapel composition.

Several creative choices were made during the implementation phase. Until the installation in the space it was not yet obvious if we wanted the vocal sounds to be fed back into the headphones during the cocoon experience. We ended up opting for no audio feedback for three reasons. First we were worried about reducing the musical aspect of the experience; secondly we thought people who are uncomfortable with their voice could be disturbed by their own sound; finally the experience is about the felling that a person has of his or her voice and any sound that would be feedback would feel artificial and already disconnected. Indeed, except for some singers of media professionals who expected to hear themselves as in a recording studio, the majority of people acknowledged very positively the fact of not hearing themselves. And people reported to have liked feeling connected only through tactile sensations.

Relevant statistics

From meditative, to explorative, the experience is always personal and most people reacted strongly to this novel access to their own voice. To understand the potential impact of this type of vocal practice we collected data on a subset of 36 participants. The details will be given in the next section.

Participants talked about their experience during the subsequent interview and survey questionnaires. 83 % of the participants reacted to the experience positively or very positively, and 17 % found it neutral. 0% judged the experience to have been negative. We observed that people reacted very strongly to the personal experience, in one of two quite opposite ways. Certain people reported feeling very relaxed and that feeling their voice on long continuous sounds helped them to calm down and meditate. Other people reacted in a much more explorative way, trying different sounds guided by the tactile sense. About two thirds of the people reported having had a relaxing or meditative experience (among which 72% reported having discovered vocal sound they never produced before), while about a third reported having felt more exploratory or freeing. 83% of the population reported having felt being part of the musical experience.

Informal feedback and observations

The installation is running in Paris from March to September 2014 and will then come to Cambridge MA in November 2014. As of today the system is successfully running and receives about 100 people a day. We noticed people coming with very young children and staying up to hours in the chapel as a relaxation and meditation space. The Vocal Vibrations installation gives an example of how some new type of reflection of the voice can be applied to a large-scale project experienced by thousands of people. The implementation makes the experience accessible to novices as well as professionals. The entry door to it is very simple: one only has to hold a single note, though there is a very large spectrum of sound and vocal textures to experience. Because the installation is free and open 4 days a week, people from very different cultural and social backgrounds came to try the adventure. During the two weeks spent there at the opening, we had the opportunity to hear yoga teachers, young children, senior citizens, professional singers, employers, etc reporting that those few minutes gave them an intimate, while still objective, access to the voice.

3.5 - Measurements of Vocal Practices

In this project, we conducted an experiment to investigate how new types of vocal practices can affect psychophysical activity. We know that health can influence the voice, but can a certain use of the voice influence health through modification of mental and physical state? This study took place in the setting of the Vocal Vibrations installation in March 2014 in Paris. The approach taken in this study is to let participants engage in the multisensory vocal exercise with a limited set of guidance to obtain a wide spectrum of vocal performances, then to compare characteristics of those performances to the influence it has on people.

In this project we obtain significant results suggesting that we can correlate the physiological state of people with the characteristic of their vocal practice if we also take into account biographical information, and in particular measurement of how much people like their own voice.

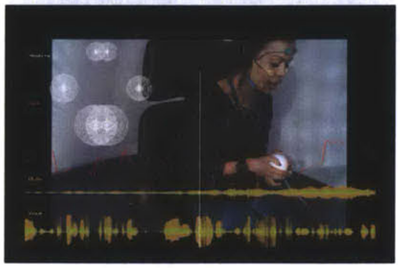


Figure 38: Synchronised view of the bioacoustic and psychophysical signals during the vocal performance of a participant

3.5.1 – Introduction, thoughts and background

After the public success of the Vocal Vibrations installation in Paris and observing how strongly people seemed to find benefits from those 6 minutes, we decided to conduct a study to verify whether we can find significant correlation between people's performance and what they experienced physiologically. In addition, the survey and interview helped us collect background and biographical information about each participant that were then compared with the effects of the performance on their psychophysiological state.

The experiment was executed by comparing people's performance during the experience to measures of their psychophysical state. To characterize people's psychophysical activity, we measure three signals: the electro dermal activity (EDA), heart rate and breathing rate. To characterize the parameters of people's vocal performance, we use the hexauscultation sensor described in chap 3.4. as well as audio vocal parameters.

Many studies have shown that meditation training (especially mindfulness meditation) may be an effective component in treating various disorders such as stress, anxiety, and chronic pain[167]. Despite the voice being a major part of several meditation traditions, the effects of the voice in meditation are mostly unexplored. In one study, medical imaging has shown that cerebral blood flow changes during meditation that incorporates chanting on resonant tones, in ways that cannot be explained solely by breath control[168].

In our case, and in concert with the Vocal Vibrations project we were

[167] Jon Kabat-Zinn, Leslie Lipworth, and Robert Burney. The clinical use of mindfulness meditation for the self-regulation of chronic pain. *Journal of behavioral medicine*, 8(2):163–190, 1985; and Richard J Davidson, Jon Kabat-Zinn, Jessica Schumacher, Melissa Rosenkranz, Daniel Muller, Saki F Santorelli, Ferris Urbanowski, Anne Harrington, Katherine Bonus, and John F Sheridan. Alterations in brain and immune function produced by mindfulness meditation. *Psychosomatic medicine*, 65(4):564–570, 2003

[168] Bangalore G Kalyani, Ganesan Venkatasubramanian, Rashmi Arasappa, Naren P Rao, Sunil V Kalmady, Rishikesh V Behere, Hariprasad Rao, Mandapati K Vasudev, and Bangalore N Gangadhar. Neurohemodynamic correlates of 'om'chanting: a pilot functional magnetic resonance imaging study. *International journal of yoga*, 4 (1):3, 2011

interested in understanding whether a certain type of vocal practice can influence people's relaxation and attention state.


Figure 39: Setting of the study

3.5.2 – Experiment design

Participants

Participants were 36 adults, between 21 and 83 years old, who came to visit the Vocal Vibrations installation at Le Laboratoire and who signed up to take part in the study after receiving all the necessary information. The study was organized over 5 days, in which we measured respectively 1, 3, 6, 13 and 13 participants. We also effected the measurements on three control participants to assess the effect on the electro-dermal activity (EDA) of staying passively in the environment setting during 6 minutes. The settings were identical throughout the 5 days in terms of music loudness, ORB settings and lighting. The study was performed while the installation was closed for the public. We took care that no visitor or employee entered the cocoon or chapel while the experience was taking place in order to limit ambient noise.

Introduction speech

Each participant, after signing the consent form, is greeted with the following introduction speech:

- Thank you very much for taking part in this study,

- The purpose of this study is to investigate how vocal sounds, vocal vibrations, singing experience and psychophysiological signals interact.

- You will have a 6:24 minutes solo experience

- In the regular Vocal Vibrations experience, you wear a microphone, a headphone with which you will hear a 6:24 minutes musical composition written by Tod Machover (as in the chapel, the music is based on the note D). The music will bring you in a journey of exploration of different sound and vocal textures.

- The idea is to really explore and play with your own voice (so please use your voice). If you can, you might want to find the D because this note will always be in harmony with the music. And from this note you can explore on different vowels, textures, sounds, etc.

- You will hold the ORB in your two hands; it will vibrate in reaction to your voice.

- For this study, in addition to this regular experience, you will be asked to wear 4 types of sensors, it will not hurt, it may be a bit disconcerting at the beginning but the goal is for you to no be distracted by them and to focus on the vocal, musical and tactile experience.

a) Maybe the most unusual of the sensors is an auscultation mask to sense the vibrations on your face (when you wear it, you can still feel free

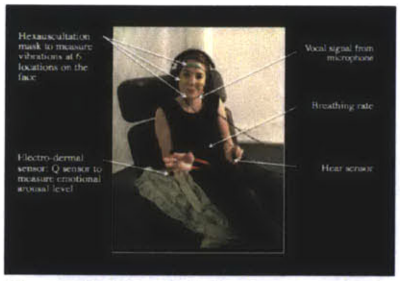


Figure 40: List of all the sensors used for the srudy

to move and articulate the sounds freely. The sensor will measure how much you make the vibrations travel in your head)

b) On your left hand you will wear an optical heart rate sensor

c) On your right hand you will wear an EDA sensor, applied with some gel

d) You will wear a belt to measure your breathing rate.

While all those sensors are designed to be comfortable and unobtrusive, you may find them irritating. If so, you should take them off and give them to the experimenter.

- You will be isolated, you should not hesitate to sing and vocalize loud if you feel like it, no-one will be disturb

- As part of this experiment you may be videotaped, you are offered three options for your confidentiality, you can

*not agree to be videotaped

*agree to allow the video to be used for this research for statistical purposes

*agree to allow the video, photo, and audio recordings to be used for this research and for educational and related non-profit purposes such as conference publications and presentations.

- After the experience, you will be asked to attend a 15 minutes interview session during which you may be asked some questions about any perceived or felt emotions throughout the experience.

- Before leaving, you will be asked to answer a brief survey questionnaire

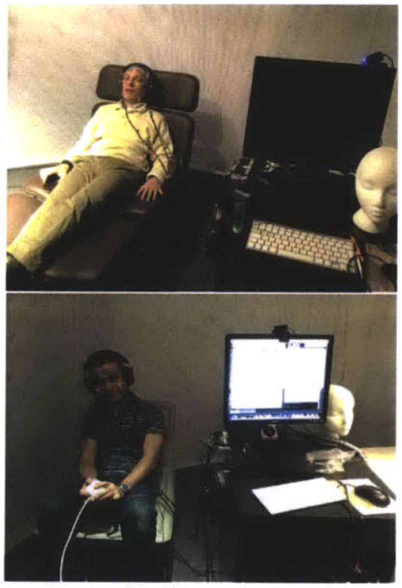


Figure 41: Participants in the setting of the study experiment

Interview

Right after the experience, the participant had a 15 minutes open ended interview during which they were still wearing the sensors. They were asked to talk about their experience through 5 questions. 1) *How did that go?* (if participant only answered with one adjective, the experimenter would follow by: can you explain specifically what you found amazing/funny/strange?). 2) *Can you recall other experiences you had that were similar to this?* 3) *At the two-thirds point of the experience (4th minute) there was a musical climax. Do you think you had a specific reaction at that moment?* 4) *Do you think this experience might have taught you something about yourself or about the vibrations produced in you body?* 5) *Do you think this kind of experience could become a practice, and if yes, what form could it take?*

Survey

In the literature, it is common to have participants answer questionnaires as soon as possible after the experience, while the memories are still fresh. We start the questionnaire with questions directly related to how they felt during the experience as well as the self-report on their level arousal and valence. We reserve biographical questions for the end. The survey questionnaire is organized in three parts: the experience, the ORB and biographical data.

Measurement system and synchronization

Although the measurement setting is done with one computer (called measurement computer MC) we also used the ORB and Analysis computers to collect data. In this part we will distinguish the three machines: measurement computer MC, analysis computer AC and ORB computer OC. To record all the data signals, we used three distinct but synchronized systems. First, as explained previously, the EDA measurement is done separately in the wearable sensor, which is synchronized with MC. Secondly we use reaper64 to record sound and vibration inputs. The setting records 8 tracks: 7 mono audio inputs and one MIDI input. The mono audio inputs are the 6 signals coming from the hexauscultation sensor plus the raw voice coming from the earset microphone. As the microphone signal goes in the AC to control the ORB behavior, we connect the audio interfaces from those two machines to read the direct voice input in MC. The MIDI input is used for time synchronization as explained later. The third system is a Java application that receives inputs from an Arduino UNO connected to the heart rate sensor and the breath sensor and from an open sound control (OSC) signal for time synchronization.

In order to know the precise start and end time of the experience, our data collection system receives an OSC message from OC at the beginning and the end of the experience. This signal is sent both to the reaper measurement system as a MIDI input and to the Java application as a timestamp.

3.5.3 - Preprocessing and interpretation of the data

VIDEO: Each video is manually analyzed according to some criteria. The motion of the hand and the body were observed to see if any motion artifacts might have occurred in the EDA signal. The level of self-consciousness of people is assessed from clues such as whether they laugh or look at the camera. An observed level of involvement and engagement is assessed from clues such as their posture on the chair, whether they keep their eyes closed or open, if they look at the camera, if they hold the ORB on their knees or keep it closer to their face, whether they put the ORB in contact to their body (chest, arms, face, etc), whether they cradle the ORB, etc.

HEARTBEAT AND RESPIRATION: The heart rate and respiration signals were not consistent and reliable enough to be used for comparison. Though the sensors and Arduino code were tested beforehand, they gave very noizy and unusable result in the real settings. We think we miscalculated the CPU power needed for video recording and this afected the reliability of the computer for receiving accurate data from the Arduino board. From the third day of experiment we decided to record the video on a different computer and the heartrate data became much more reliable. But because only fifty percent of the participants had their heartrate and breathing rate measured, we decided to not use those data for analysis.

INTERVIEW: The answers from interview questions give us insight into the overall enthusiasm of people and help us understand the case-by-case individual experience of participants.

SURVEY QUESTIONS, PROCESSING AND INTERPRETATION: To remove individual bias from the survey, we rescale the results of all numerical questions between 0 and 5 for each participant individually by a applying feature scaling. If X is the answer to a question and Xmax and Xmin are the maximum and minimum scores given by the same participant on the whole survey, then we consider X' the scalled results given by $X' = 1 + \frac{(5-1)*(X-Xmin)}{Xmax-Xmin}$

PROCESSING OF THE VOICE SIGNAL: We use Praat, a software specialized for phonetic analysis of the voice[169]to extract characteristics of the vocal performance. The data we extract are: Median pitch (Hz) , Mean pitch (Hz), Standard deviation (Hz), Minimum pitch (Hz), Maximum pitch (Hz), Fraction of vocally unvoiced frames, Jitter, Shimmer, Mean autocorrelation, Mean noise-to-harmonics ratio and Mean harmonics to noise ratio (dB). To measure whether people were staying on the note D, we defined a Dness parameter. The note D was chosen as core of the music partially because this note is often present in two octaves in people's vocal range. Indeed many participants were jumping from one octave to another at different moments of the experience. In consequence, the value of mean pitch or median pitch is not adapted to assess if people stayed close to the D. In order to measure the Dness parameter, we track the pitch of the overall experience while imposing a pitch value present in one octave only.

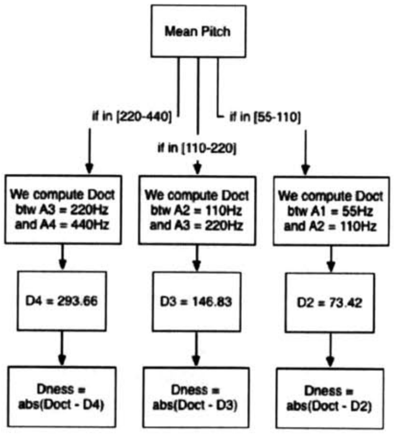


Figure 42: Method used to compute the Dness parameter

EDA, MEASUREMENT AND INTERPRETATION: To measure the participant's EDA, we used an affectiva Q sensor. We preliminarily set up the time synchronization with the machine used for other measurements to simplify data synchronization. We had set up the data sample rate to 8Hz. We placed the sensors at the participant's fingertips rather than on their wrist, where the signal is weaker. We used medical gel on the fingers to improve the signal reading and we attached the electrodes with Velcro on the little finger and ring finger of the right hand. Once the sensor is in place it needs some time for the signal to reach the right level. We took care to install the EDA sensor 5 minutes before the beginning of the experience.

---

[169] P Boersma and D Weenink. Praat manual: doing phonetics by computer (v. 5.0. 23).[computer program], 2008

During that time, the experimenter was helping the participant to install the rest of the sensors and to give the last set of instructions.

Figure 43: EDA Q sensors from the Affectiva company

To preprocess the EDA signal, we extract a time stamped version of the signal from the Q software. Then we import in Matlab and crop the 6:24 minutes corresponding to the experiment. The format is then modified to fit the format requiered by LedaLab ((removing part of the front matter and aligning the result into the right CVS format) using code written and graciously shared by Javier Hernandez. Then the signal is separated between phasic and tonic signal using the LedaLab Matlab extension[170]. The tonic part of the signal corresponds to the background variation while the phasic part is the rapid component resulting from sympathetic neuronal activity. From this, we extract the number of peaks above the median threshold in the phasic signal. We watched the video of the experience to remove potential motion artifacts.

Because everyone has a different EDA level and reactivity we need an individual baseline. We extract this baseline from two minutes of the interview time, when people answer to the first interview question: *how did it go?* This baseline is not perfect and the result might be influenced by the fact that the people are asked to speak, to remember their experience and to interact to the experimenter. But we still consider that it gives a good enough piece of information about the individual variation of EDA from each participant outside the experimental context and to know whether certain participants are likely to be hyper- or hypo-responders[171].

From those considerations we define a certain number of metrics that will be used to assess the results. EDA_N_A (aligned normalized EDA) is the raw EDA signal normalized over the experience and from which we subtract the initial value in order to align the signal to the initial state to see the evolution trend brought by the experience. EDAT_N_A (aligned normalized tonic EDA) is the tonic component of the EDA passed through the same process. EDAP_N (normalized phasic EDA) is the phasic component of the EDA normalized throughout the experiment. EDA_peaks is the number of peaks above the median threshold from EDAP_N normalized by the number of peaks measured the same way on the participant's baseline signal. For the 36 participants, the EDA_peaks value goes from 8 to 140. The higher the number, the higher the arousal level was during the interaction

One has to be careful when trying to interpret EDA. It has been proven that it is often linked with sympathetic nervous system arousal and so can give insight about the arousal level of people[172]. But one challenge comes from the specificity problem described in Elliott Hedman PhD's dissertation as follows: "We may be unable to determine what factor(s) resulted in a specific physiological response[173]." Arousal is a measure

---

[170] Mathias Benedek and Christian Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, 190(1):80–91, 2010a; and Mathias Benedek and Christian Kaernbach. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4):647–658, 2010b

[171] B Iopac Mpr, Acq Knowledge, Robert Jones, and Mickey Rowe. A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. pages 1–42, 2013

[172] Hugo D Critchley. Book review: electrodermal responses: what happens in the brain. *The Neuroscientist*, 8(2):132–142, 2002; and Elliott Hedman, Oliver Wilder-Smith, Matthew S Goodwin, Ming-Zher Poh, Rich Fletcher, and Rosalind Picard. icalm: Measuring electrodermal activity in almost any setting. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–2. IEEE, 2009

[173] Elliott Hedman. *Observational Psychophysiology for Empathic Design*. PhD thesis, Massachusetts Institute of Technology, 2014

of how responsive people are to stimuli; it involves the activation of the reticular activating system (RSA) and the endocrine system. To understand the psychophysiological meaning of arousal, it is common to couple it with the measure of valence. Figure 44 shows the schematic for the two-dimensional structure of affect adapted from[174]. The knowledge of valence enables to label arousal values to emotional states. Indeed, in our study, 100 percent of the people reported a positive valence after the experience, which we could interpret as: peaks in the EDA signals do not correspond to stress but to high level of engagement, while a low and decreasing EDA would not correspond to boredom but to a calm state or relaxation. In our case, understanding the EDA value as "how responsive people are to stimuli" is as relevant as linking it to specific emotions. Indeed, one possible objective is to focus on the task as a self-learning task and use the EDA as a measure of flow[175].

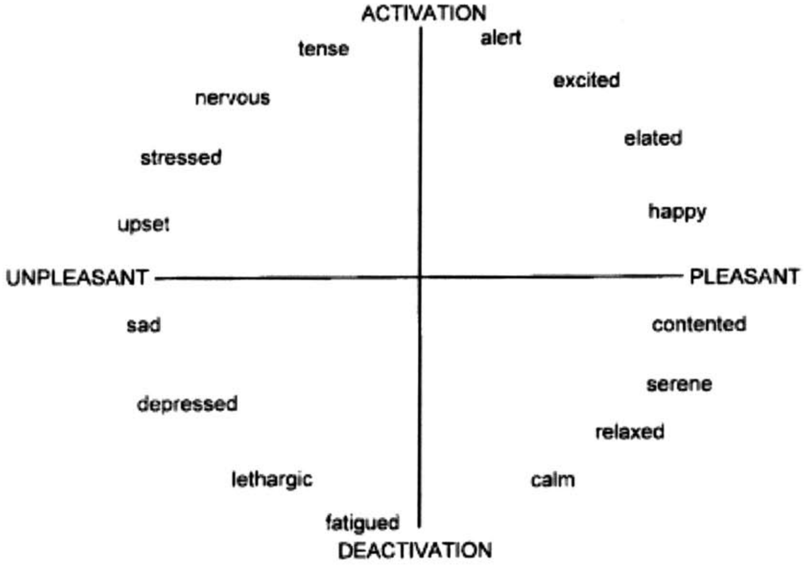


Figure 44: Schematic for the two-dimensional structure of affect. Adapted form Feldman Barnett and Russel (1998)

Interpretation of the Vibration signal: we took the same approach as in research measuring the level of nasalisation by computing the ratio between amplitude signals measured at the nose and neck skin[176]. In our case we measure to what extent people were exploring different vibration patterns as opposed to staying on the same sound during whole the experience. Figure 45 shows the key steps of the process.

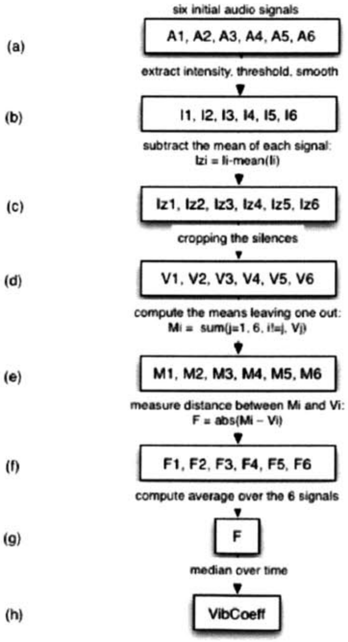


Figure 45: Method used to compute the VibCoeff parameter

The six signals from the Hexauscultation device have to be preprocessed in order to be compared. From the 6 audio measures, we extract the intensity levels using the Praat software (step (a)) and we threshold and smooth the signal by moving average with a span of 5 (step (b)). Each signal

[174] Lisa Feldman Barrett and James A Russell. Independence and bipolarity in the structure of current affect. *Journal of personality and social psychology*, 74(4):967, 1998

[175] Corinna Peifer. Psychophysiological correlates of flow-experience. In *Advances in flow research*, pages 139–164. Springer, 2012

[176] RP Lippmann. Detecting nasalization using a low-cost miniature accelerometer. *J Speech Hear Res*, 24(3):314–317, 1981; Margaret Redenbaugh. Correspondence between an accelerometric nasal/voice amplitude ratio and listeners' direct magnitude estimations of hypernasality. *Journal of speech and hearing research*, 28(2):273, 1985; and Yoshiyuki Horii. An accelerometric measure as a physical correlate of perceived hypernasality in speech. *Journal of speech and hearing research*, 26(3):476, 1983

is then detrended (we subtract the mean value of the time-series signal) resulting in a zero-mean signal (step (c)). The result of the comparison should not be influenced by the voiced percentage of the experience: all silent moments would show a very high correlation between the signals. To avoid this, we only crop the voiced signal and do the comparison on those (step (d)).

To compare them, one possible method would be to then compute the mean intensity signal and for each time sample, then to compute the variance of the vector [V0(t), V1(t), V2(t), V3(t), V4(t), V5(t), V6(t)], and to use the average of this value over time as a measure of vibration richness. In our case, we were more interested in how much the signals change relative position. Figure 46 shows the behaviour of the signals we want to highlight: the less correlated the 6 signals are, teh more we consider the vibrations patern to be rich. To measure this we use a Jackknife method. We compute 6 times the mean of the signals leaving one out (step (e)). Mi = sum(j=1, 6, i!=j, Vj). Then we measure the distance between each signal and its associated mean F(t) = abs(Mi(t) – Vi(t)) (step (f)).

We compute the average of the 6 distance signals (step (g)) and we use the median of this time-signal as a measure of vibration pattern richness that we call VibCoeff (step (h)). We normalize the result between 0 and 10. Figure 46 and 47 show an example of a pattern of vibrations with a high VibCoeff and one with a low VibCoeff over time.

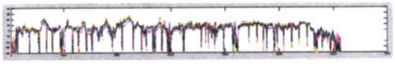


Figure 46: Example of vibrations pattern with a low VibCoeff = 2.58

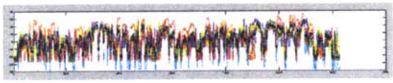


Figure 47: Example of vibrations pattern with a HIGH VibCoeff = 8.08

3.5.4 – Results

The objective of this study is to see whether the performance characteristics have an implication for the EDA results. In our case the null hypothesis can be phrased as: H0 = there is no link between the EDA_peaks and VibCoeff.

We judge the validity of our hypothesis through the p-value that gives us the probability of observing the given sample result under the assumption that the null hypothesis is true. When observing the relation between EDA_peaks and VibCoeff throughout the whole population, no trend can be extracted. Figure 48 shows the linear regression curve for all the participants. On the graph, no pattern is apparent. The results of this regression are shown in Figure 49

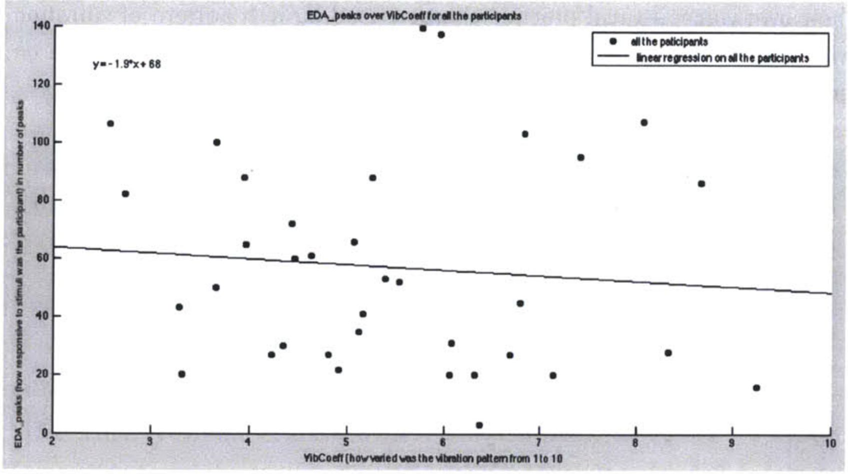
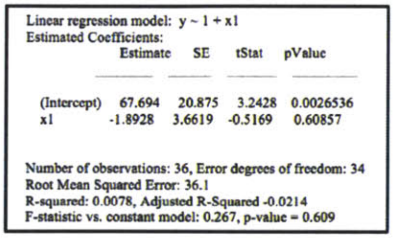


Figure 48: Plot of the linear regression for all the participants



```
Linear regression model: y ~ 1 + x1
Estimated Coefficients:
                  Estimate     SE      tStat    pValue
                  ________   ______   _______  _________

    (Intercept)    67.694    20.875    3.2428  0.0026536
    x1            -1.8928    3.6619   -0.5169   0.60857

Number of observations: 36, Error degrees of freedom: 34
Root Mean Squared Error: 36.1
R-squared: 0.0078, Adjusted R-Squared -0.0214
F-statistic vs. constant model: 0.267, p-value = 0.609
```

Figure 49: Table showing the results of the linear regression for all the participants

The p-value we obtain is very high (>0.1) which means there is no presumption against the null hypothesis according to statistical significance testing.

However when the population is divided in subgroups corresponding to the survey answer to the question "how much do you like your voice" the results become more interesting. We call group A the people who answered: "I like my voice quite a lot," B the group of people who answered: "I like it a bit" and C the group who answered "I do not like it at all." Figure 50 shows the linear regression curve on group A. The results of this regression are shown in Figure 51.

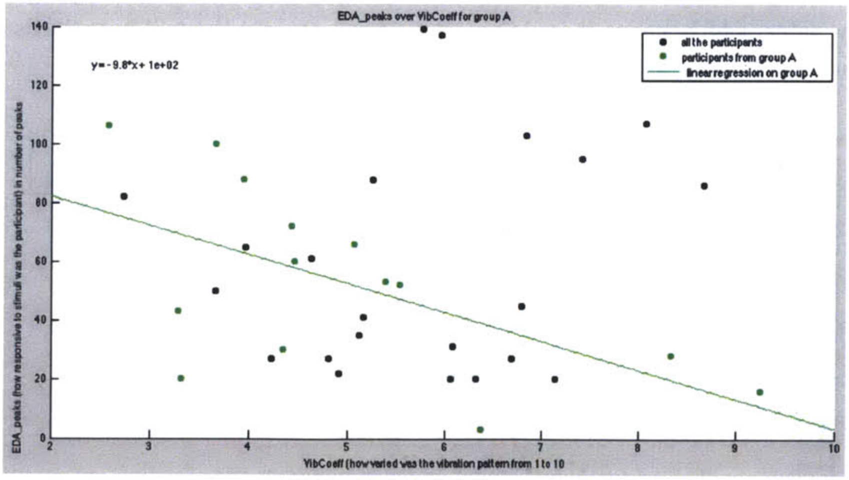
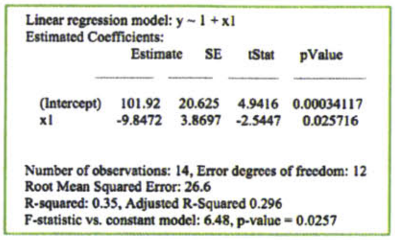


Figure 50: Plot of the linear regression for the participants who answered that they like their own voice: group A



```
Linear regression model: y ~ 1 + x1
Estimated Coefficients:
                  Estimate     SE      tStat     pValue
                  ________   ______   _______  __________

    (Intercept)    101.92    20.625    4.9416  0.00034117
    x1            -9.8472    3.8697   -2.5447   0.025716

Number of observations: 14, Error degrees of freedom: 12
Root Mean Squared Error: 26.6
R-squared: 0.35, Adjusted R-Squared 0.296
F-statistic vs. constant model: 6.48, p-value = 0.0257
```

Figure 51: Table showing the results of the linear regression for the participants who answered that they like their own voice: group A

The p-value we obtain is <0.05 which gives us a strong presumption against the null hypothesis. According to statistical significance testing we consider this result as significant and suggest that for people who "like" their own voice, a vocal practice characterised by rich pattern of vibration will result in low arousal. Figure 46 shows the linear regression curve on group C. The results of this regression are shown in Figure 55

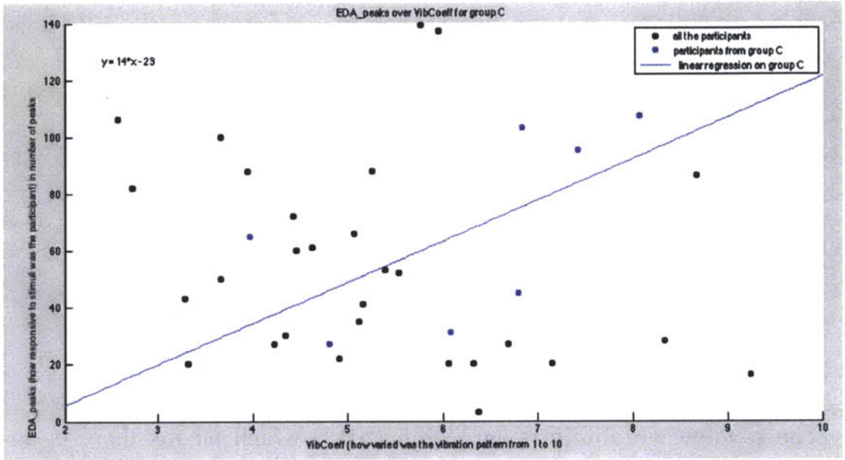
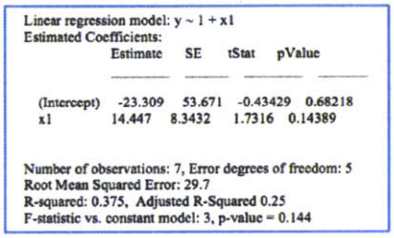


Figure 52: Plot of the linear regression for the participants who answered that they do not like their own voice at all: group C



```
Linear regression model: y ~ 1 + x1
Estimated Coefficients:
                 Estimate    SE       tStat     pValue
                 ________   ______   ________   _______

  (Intercept)    -23.309    53.671   -0.43429   0.68218
  x1              14.447    8.3432    1.7316    0.14389

Number of observations: 7, Error degrees of freedom: 5
Root Mean Squared Error: 29.7
R-squared: 0.375,  Adjusted R-Squared 0.25
F-statistic vs. constant model: 3, p-value = 0.144
```

Figure 53: Table showing the results of the linear regression for the participants who answered that they do not like their own voice at all: group C

The p-value we obtain is slightly >0.1 but is still inferior to the value obtained for all the participants together. We can suggest that for people who are not at all at ease with their own voice, a vocal practice characterized by a rich vibration pattern will result in higher arousal. The results have to be tempered by the fact that this group contains only seven participants.

Finally for group B, the result from linear regression shows a very high p- value, which means there is no presumption against the null hypothesis according to statistical significance testing.

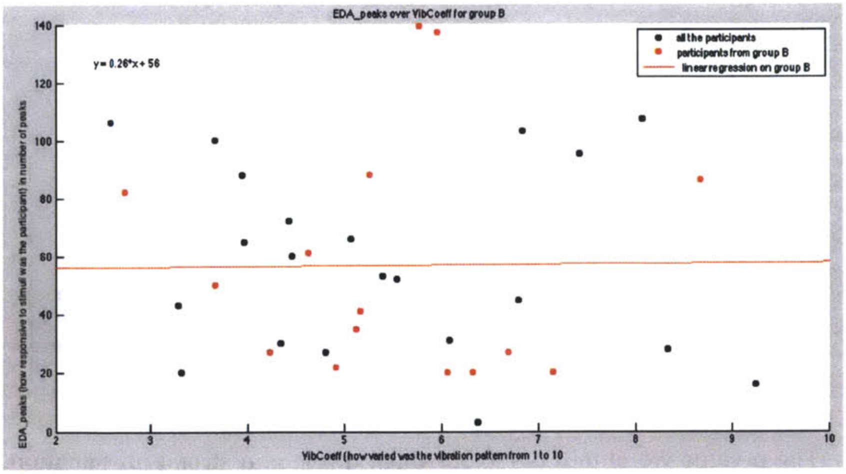
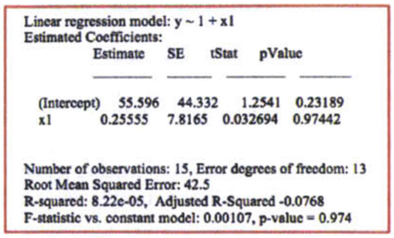


Figure 54: Plot of the linear regression for the participants who answered that they liked their voice a bit: group B



```
Linear regression model: y ~ 1 + x1
Estimated Coefficients:
                 Estimate    SE       tStat      pValue
                 ________   ______   ________   _______

  (Intercept)    55.596     44.332    1.2541    0.23189
  x1             0.25555    7.8165   0.032694   0.97442

Number of observations: 15, Error degrees of freedom: 13
Root Mean Squared Error: 42.5
R-squared: 8.22e-05,  Adjusted R-Squared -0.0768
F-statistic vs. constant model: 0.00107, p-value = 0.974
```

Figure 55: Table showing the results of the linear regression for the participants who answered that they do not like their own voice at all: group C

3.5.5 - Discussion

Statistical analysis

The question that motivated this study was whether we could significantly show that vocal practices could act on psychophysical activity. Our experimental results suggest that given certain knowledge on biographical data we can predict how people would react to a certain set of vocal exercises. By letting people be relatively free in their vocal experience we have been able to assess the performances according to certain criteria of vibration patterns. We have described how the correlation of these practices and biographical data can predict the influence that certain practices can have on people's state of relaxation or engagement. This study would require further experimentation to verify that imposing a specific vocal practice would not bring other perturbations.

Our results suggest that for people who have an harmonious relationship with their own voice, rich vibration patterns lead to low EDA, which can be interpreted as meaning that the more they focus on the exercise and contemplate their voice as an instrument through the ORB, the more they will escape form their normal condition and be driven into a relaxing and meditative state. When people from group A tried various vibrations patterns, it helped them calm down, "disactivate" and put their mind at rest. When people from group A stayed on a static vibration pattern, the exercise didn't help them to be very focused and calm.

For the group C composed otf people who are not very confident with their voice, on the other hand, the exercise consisting of trying on one sound and keeping it during all the experience seems to help them relax more than if they try very rich sound patterns.

Those results suggest that we can correlate the physiological state with the characteristic of their vocal practice if we also take into account biographical information, and in particular measurement of how much people like their own voice. This study is a step toward designing Vocal Self Reflective Experiences that are personalised and designed to lead people into certain chosen states (meditation, relaxation or exploration).

Global observations

The survey and interview results showed that people found the experience very positive. In the survey 83% of the participants answered that they reacted positively or very positively to the experience and 72% reported having discovered vocal sound they never produced before. During the interview, 80.5% of the participants answered that they could imagine this type of experience becoming a practice they would want to do again.

Among the people who consented having their performance recorded, 26% laughed at least once during the interaction and 15% had at least a glance at the camera. 52% kept their eyes closed while the other 48% were looking at the orb. We can suggest that people felt quite connected with the ORB, indeed 63% were cradling the ORB during the interaction and 37% of the people connected it to their body (mainly neck, throat and belly). In the survey, 83% of the participants reported having felt connected or very connected with the ORB.

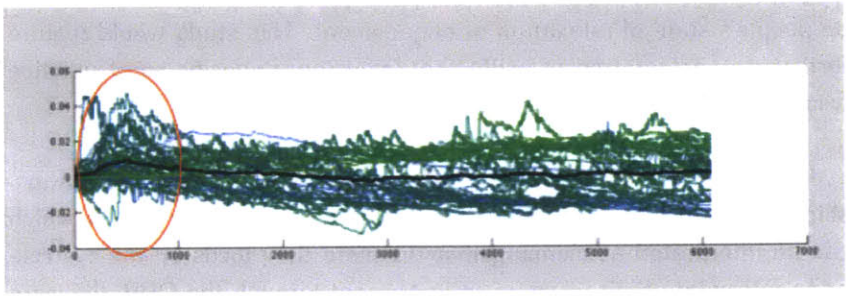


Figure 56: EDA bump in the first 30 seconds the green curves represent the 36 EDA_N_A signals of participants, the black curve correspond to the mean signal of the 36 curves. The signals go negative values because the EDA signals underwent a substraction by their mean. The x axis is in seconds

On the EDA signal, we systematically observe a "bump" in the first 30 seconds of the experience. From people's comments in the interview we can suggest that it comes from the first contact with the ORB's vibrations. People are not insensitive to this very new way to connect with their voice and we observe this systematic reaction from people who have this experience for the first time. This hypothesis should be moderated: maybe this bump is the result of people expecting something, and they just got left alone in the room and have the excitement of listening to the music starting. Figure 56 represents the 36 EDA_N_A signals of participants. The black curve correspond to the mean of the 36 curves.

Interview excerpts

The oral answers from the interview gave us a wealth of information, suggestions and descriptions of experience. When asked how the experience went, we collected answers such as:

- "I had lots of fun!"
- "It is a bit intimidating at the beginning, but there is a real joy to follow, and to let go to how your being wants to sound"
- "I felt completely somewhere else, I could have kept doing that for 30 minutes"
- "I really wanted to make it touch my body, belly or throat"
- "It went well"

- "There are moments when we really don't know what we are doing"
- "Feeling the vibrations from one part of the object to another, also switching conscious focus between my own voice and the music"
- "Having the object is really great because it makes the voice tangible in a different way"
- "I feel much more alert than 6 min ago"

When asked to describe analog situations, people reported:

- " I adored it! I felt like I was Ariel the little mermaid who sees her voice going out of her body"
- "Like a form of meditation"
- "My little boy is 2 years old and I sometimes sing lullabies for him that are self-composed and based on the feeling of the moment, it reminded me of that"
- "[It reminded me of] being in mass, in chuch, or singing in a choir"
- "Like the final moments"
- "Like all first times, this feeling of discovering something completely new, never seen before"

When asked if people could see this type of exercise becoming a practice, we collected answers such as:

- "Good way to be put in front of yourself"
- "To think about nothing, except our voice, ourselves and forgetting the problems we have, like yoga maybe."
- "The orb really made me want to prove to myself all the different things I can do, and how I can let it go"
- "Tt can help enter into meditation or just for relaxing or just for playing."
- "It felt like my voice was tensed when I sang with 'thick cords technique' I could feel the tensions in the ORB, when I sang with a 'thin cords technique', it felt much lighter in the ORB."
- "I would like to have such a tool when I work with people for voice reeducation, to develop a link between you and yourself but passing through this [the ORB], to help people develop peoples' feelings"
- "It re-appropriates your own body, I would go even further, when we talk about re-appropriating the voice, so the vibration, so the motion of our body, putting the body in motion it makes me think of people who are fixed and frozen, people who have lived traumas, and helping them first with their voice because the voice is a motion of the body. Because when people live traumas they sometimes disembody themselves and this could really help"
- "It is something that people don't really allow themselves to do, playing with their own voice. who cares if it is not very pretty it really helps make us feel better."

4 - Conclusion

4.1 - Summary

In this thesis, we explored the design of experiences that help people learn about themselves through their voice. We defined the space of Self-Reflective Vocal Experiences as mirror-like installations, situations or experiences designed to encourage users into changing their perception and exploring the connection they have with their voice, and beyond this, the experience they have with themselves.

We presented a cross-referenced reading of many domains centered on the voice to extract personal elements that the voice reveals. Different aspects of the voice have been studied in different domains from clinical medicine, biology and surgery to psychology, physics, linguistics, communication theory, learning sciences, etc. Our work proposed methods and examples on how to use knowledge from those different fields to make them accessible and experienciable in meaningful ways in a self-reflective context. In addition, we also offered a history of projects and installations that embrace the characteristics of self-reflective vocal systems.

We then presented the journey of thoughts followed by the author in the quest of different modes of self-reflection accessible through the voice: the flow of our explorations started from a very physical, embodied and extroverted type of self-reflection; followed by a more social and networked type of self-reflection through our interaction with others when participating in the human social experience; finally to arrive at a more personal and introverted type of self-reflection.

Finally we presented four new systems that explore the properties of a Self-Reflective Vocal Experience inspired by the bibliographical research as well as the preliminary explorations. Each project targets a specific aspect of voice to improve our understanding of it. The *Voice minus Words* project offers an experience to shift people's attention from the words said to how the words are said and gives access to the speech normalized from the words to enable people to focus on what remains. The *Vocal*

Vibrations Visualization project offers a real-time visual representation of the amplitude of vocal vibrations at six locations on the face. It is designed to help gain better control of the voice as a physical embedded instrument. The *Vocal Vibrations* installation offers a rich and active musical vocal experience to engage the public in the regular practice of thoughtful singing and vocalizing, both as an individual experience and as part of a community. In the context of the installation, the *ORB* helps people gain awareness of the physicality of their voice through vibrations. Finally, the project called *Measurement of Vocal Practices* presents significant results suggesting that we can correlate the physiological state of people with the characteristic of their vocal practice if we also take into account biographical information, and in particular measurement of how much people like their own voice.

4.2 - Conclusions and Future Directions

Vocal Practices

In the last project presented, we observed that for people who like their voice (group A) the more rich the vibration pattern is, the more they enter into relaxation state. The trend is opposite for people who say they don't like their voice (group C). This study is a step toward designing Vocal Self Reflective Experiences that are personalized and designed to lead people into certain chosen states (meditation, relaxation or exploration). This result supports the values of placing the relationship people have with their own voice at the core of the experience design. In our future research path, we want to look at this aspect both as a design parameter and an objective.

For the first category of investigation (voice as a design parameter), we want to further the understanding of how framing and perception of the voice can modify someone's behavior. For the second category of experimentations (voice as an objective), we wish to help people regain self-confidence in their voice and thus work on reconnecting their persona and their self. We want to work on the development of practices that take into consideration people's individual characteristics and development to improve their mental and physical state. Our objective is not to change people's voice to fit a model but to show that everyone has the option of thought when using their voices. We want to design more systems to help people tune to themselves, reconnect with their own body and their past through an acceptance and an enjoyment of their voice. We also want to explore more in depth the notion of free will in the use of the voice and maybe use this as an entry to general human free will .

Role of Technology

Technology is not essential in creating Self-Reflective Vocal Experiences. Indeed one very simple experience could consist of placing one hand behind your ear as a shell/receptor, and the other hand in front of your mouth to conduct the sound and create a tunnel from your mouth to your ear. If you then produce vocal sound, you will perceive your voice in a very different way, filtered by the waveguide created by your hands.

In our designs, we use the concept of estrangement to help people change perspective about their everyday life vocal activity. In this work we reversed the common connotations associated with the word "estrangement" and we define it as the freeing act of taking distance and looking at the world with "the eyes of a horse or a child."[177] This is the act of making "foreign" something that has become so familiar that we don't even see it anymore. If technology is not essential to pursue our goal, it can become a very useful tool to help us enrich our perspectives.

The methods proposed in this thesis also present a unique direction in self-reflection and vocal discovery in that they engage the user to perceive herself and what she projects of herself in social context and affective interactions. Because humans tend to project a lot of themselves onto the technology they interact with, the technology can actively be used as a mirror of themselves. Here this aspect is highlighted and the goal is not to make the technology transparent but to make it a frame to embed some part of the user's personality. To that extent, one idea behind this thesis is to promote the use of technology as a mirror to design interactions between the users and parts of themselves that they rarely get access to.

We hope that the cross-disciplinary approach demonstrated in this work is just the beginning of this new way of considering the voice and its relation with our physical, social or mental identity. Also we hope this work will help connect the fields and offer scientists and designers the first steps of a common language to work together on products and experience to improve people's quality of life. We believe in an approach that helps people tell their own story using their voice as a medium, a content and a material, becoming mindful of where they come from and what their roots are. We believe in the value of providing them not only with a way to use their voice in a thoughtful way to assume and remember who they are, but also with a lens to learn the value of the aesthetic dimension of life.

---

[177] Carlo Ginzburg. À distance. Neuf essais sur le point de vue en histoire. 1998

Bibliography

La laboratoire. URL `http://www.lelaboratoire.org/en/archives-18.php`.

Vocal vibrations. URL `http://opera.media.mit.edu/projects/vv/`.

Interview in documentary "listen", 1992.

David A. Lieberman Wendy L. Sunnucks Jamie DJ Kirk. Reinforcement without awareness: I. voice level. *The Quarterly Journal of Experimental Psychology: Section B*, 51(4):301–316, 1998.

Mariko Akutsu and Yasuhiro Oikawa. Extraction of sound field information from flowing dust captured with high-speed camera. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 545–548. IEEE, 2012.

Mariko Akutsu, Yasuhiro Oikawa, and Yoshio Yamasaki. Extract voice information using high-speed camera. 19:055019–055019, 2013.

Christos-Nikolaos Anagnostopoulos, Theodoros Iliou, and Ioannis Giannoukos. Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. *Artificial Intelligence Review*, November 2012.

Barry Arons. A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 12(7):35–50, 1992.

Jean-baptiste Barriere. Project : the From Synthesis of to Voice the Singing Synthesis in General The. 8(3):15–31, 2010.

Olivier Bau, Ivan Poupyrev, Mathieu Le Goc, Laureline Galliot, and Matthew Glisson. Revel: programming the sense of touch. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 2785–2786. ACM, 2013.

Georg V Békésy. The structure of the middle ear and the hearing of one's own voice by bone conduction. *The Journal of the Acoustical Society of America*, 21(3):217–232, 1949.

P Belin, R J Zatorre, P Lafaille, P Ahad, and B Pike. Voice-selective areas in human auditory cortex. *Nature*, 403(6767):309–12, January 2000.

Daryl J Bem. Self-perception theory1. 1973.

Mathias Benedek and Christian Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, 190(1): 80–91, 2010a.

Mathias Benedek and Christian Kaernbach. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4):647–658, 2010b.

Jonathan Berger and Song Hui Chon. Simulating the sound of one ' s own singing voice. 2003.

BM Blackwood and Henry Balfour. Ritual and secular uses of vibrating membranes as voice-disguisers. *Journal of the Anthropological Institute of Great Britain and Ireland*, pages 45–69, 1948.

P Boersma and D Weenink. Praat manual: doing phonetics by computer (v. 5.0. 23).[computer program], 2008.

Hans-Georg Bosshardt. Cognitive processing load as a determinant of stuttering: summary of a research programme. *Clinical linguistics & phonetics*, 20(5):371–85, July 2006.

M Botvinick and J Cohen. Rubber hands 'feel' touch that eyes see. *Nature*, 391(6669):756, February 1998.

Adam Boulanger. *Autism, new music technologies and cognition.* PhD thesis, Massachusetts Institute of Technology, 2006.

John Cage. The future of music: Credo. *Silence: Lectures and Writings by John Cage*, pages 3–7, 1973.

Janet E Cahn. The generation of a ect in synthesized speech. *Journal of the American Voice I/O Society*, 8:1–19, 1990.

William E Cayley Jr. Diagnosing the cause of chest pain. *Am Fam Physician*, 72(10):2012–2021, 2005.

Angela Chang, Sile O'Modhrain, Rob Jacob, Eric Gunther, and Hiroshi Ishii. Comtouch: design of a vibrotactile communication device. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 312–320. ACM, 2002.

Charles Chaplin and Paulette Goddard. *The Great Dictator: Film.* 1940.

Marie Cheour-Luhtanen, Kimmo Alho, Kimmo Sainio, Teemu Rinne, Kalevi Reinikainen, Maija Pohjavuori, Martin Renlund, O Aaltonen, O Eerola, and Risto Näätänen. The ontogenetically earliest discriminative response of the human brain. *Psychophysiology*, 33(4):478–481, 1996.

R. D. Cherry. Ethos Versus Persona: Self-Representation in Written Discourse. *Written Communication*, 5(3):251–276, July 1988.

John Chowning. "Phoné" (1980-81). 1981.

John Chowning. Fifty Years of Computer Music : Ideas of the Past Speak to the Future. 2007.

Stephen Clift. The significance of choral singing for sustaining psychological wellbeing : findings from a survey of choristers in England , Australia and Germany. pages 79–96, 2010.

Stephen Clift, Grenville Hancox, Ian Morrison, Bärbel Hess, Gunter Kreutz, and Don Stewart. Choral singing and psychological wellbeing: Findings from english choirs in a cross-national survey using the whoqol-bref. In *Proceedings of the International Symposium on Performance Science*, pages 201–207, 2007.

Amon Cohen and AD Berstein. Acoustic transmission of the respiratory system using speech stimulation. *Biomedical Engineering, IEEE Transactions on*, 38(2):126–132, 1991.

Walter Cohen. Color-Perception in the Chromatic Ganzfeld. *The American Journal of Psychology*, 71(2):390–394, 2013.

Francis Macdonald Cornford. *Plato's theory of knowledge: The theaetetus and the sophist*. Courier Dover Publications, 2003.

AD Craig. Human feelings: why are some more aware than others? *Trends in cognitive sciences*, 8(6):239–241, 2004.

Arthur D Craig. How do you feel? interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, 3(8): 655–666, 2002.

Hugo D Critchley. Book review: electrodermal responses: what happens in the brain. *The Neuroscientist*, 8(2):132–142, 2002.

Hugo D Critchley, Stefan Wiens, Pia Rotshtein, Arne Öhman, and Raymond J Dolan. Neural systems supporting interoceptive awareness. *Nature neuroscience*, 7(2):189–195, 2004.

William A Curtin. Theory of mechanical properties of ceramic-matrix composites. *Journal of the American Ceramic Society*, 74(11):2837–2845, 1991.

Cymascope. Cymascope, sound made visible. URL `http://cymascope.com/cyma_research/phonology.html`.

Dirk Cysarz, Dietrich von Bonin, Helmut Lackner, Peter Heusser, Maximilian Moser, and Henrik Bettermann. Oscillations of heart rate and respiration synchronize during poetry recitation. *American Journal of Physiology-Heart and Circulatory Physiology*, 287(2):H579–H587, 2004.

Antonio Damasio. *Descartes' error: Emotion, reason and the human brain.* Random House, 2008.

Adrian Daub. " hannah, can you hear me?"—chaplin's great dictator," schtonk," and the vicissitudes of voice. *Criticism*, 51(3):451–482, 2009.

Richard J Davidson, Jon Kabat-Zinn, Jessica Schumacher, Melissa Rosenkranz, Daniel Muller, Saki F Santorelli, Ferris Urbanowski, Anne Harrington, Katherine Bonus, and John F Sheridan. Alterations in brain and immune function produced by mindfulness meditation. *Psychosomatic medicine*, 65(4):564–570, 2003.

Ana C Andrés del Valle and Agata Opalach. The persuasive mirror. In *Proc. Int'l Conf. Persuasive Technology*, 2006.

Pierre C Delattre, Alvin M Liberman, and Franklin S Cooper. Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, 27(4):769–773, 2005.

Kelly Elizabeth Dobson. Objects that interact with a user at a visceral level, October 1 2013. US Patent 8,545,228.

DG Doehring and BN Bartholomeus. Laterality effects in voice recognition. *Neuropsychologia*, 9(4):425–430, 1971.

Judith Donath, Karrie Karahalios, and Fernanda Viegas. Visualizing conversation. *Journal of Computer-Mediated Communication*, 4(4):0–0, 1999.

J. Droppo and a. Acero. Analysis and comparison of two speech feature extraction/compensation algorithms. *IEEE Signal Processing Letters*, 12 (6):477–480, June 2005.

Paul Duchnowski, Uwe Meier, and Alex Waibel. See me, hear me: integrating automatic speech recognition and lip-reading. In *ICSLP*, volume 94, pages 547–550. Citeseer, 1994.

Anders Ekblom and Per Hansson. Extrasegmental transcutaneous electrical nerve stimulation and mechanical vibratory stimulation as compared to placebo for the relief of acute oro-facial pain. *Pain*, 23(3):223–229, 1985.

Jon Erickson. The Language of Presence : Sound Poetry and Artaud. 14 (1):279–290, 2013.

Sarah Evans, Nick Neave, and Delia Wakelin. Relationships between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice. *Biological psychology*, 72(2):160–163, 2006.

Norman AS Farb, Zindel V Segal, Helen Mayberg, Jim Bean, Deborah McKeon, Zainab Fatima, and Adam K Anderson. Attending to the present: mindfulness meditation reveals distinct neural modes of self-reference. *Social cognitive and affective neuroscience*, 2(4):313–322, 2007.

Lisa Feldman Barrett and James A Russell. Independence and bipolarity in the structure of current affect. *Journal of personality and social psychology*, 74(4):967, 1998.

Raul Fernandez. *A computational model for the automatic recognition of affect in speech*. PhD thesis, Massachusetts Institute of Technology, 2004.

William P Fifer and Christine M Moon. The role of mother's voice in the organization of brain function in the newborn. *Acta Paediatrica*, 83(s397): 86–93, 1994.

Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):5, 2002.

WD Gaillard, M Pugliese, CB Grandin, SH Braniecki, P Kondapaneni, K Hunter, B Xu, JR Petrella, L Balsamo, and G Basso. Cortical localization of reading in normal children an fmri language study. *Neurology*, 57(1): 47–54, 2001.

Shaun Gallagher. Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4(1):14–21, 2000.

Karyn L Galvin et al. A comparison of tactaid ii and tactaid 7 use by adults with a profound hearing impairment. *Ear and hearing*, 20(6):471, 1999.

Pritesh Gandhi. Ambient device company. URL `http://www.ambientdevices.com/`.

Carlo Ginzburg. À distance. Neuf essais sur le point de vue en histoire. 1998.

By Evelyn Glennie. Hearing Essay. 1993.

Mel Goldberg. Face mask with voice modifying capability, July 28 1987. US Patent 4,683,588.

Julio González, Alfonso Barros-Loscertales, Friedemann Pulvermüller, Vanessa Meseguer, Ana Sanjuán, Vicente Belloch, and César Ávila. Reading< i> cinnamon</i> activates olfactory brain regions. *Neuroimage*, 32(2):906–912, 2006.

Mel Gordon. Laurie Anderson, performance Artist. 24(2):51–54, 2013.

Christina Grape, Maria Sandgren, Lars-Olof Hansson, Mats Ericson, and Töres Theorell. Does singing promote well-being?: An empirical study of professional and amateur singers during a singing lesson. *Integrative physiological and behavioral science : the official journal of the Pavlovian Society*, 38(1):65–74, 2003.

Maria I Grigos and Rupal Patel. Articulator movement associated with the development of prosodic control in children. *Journal of speech, language, and hearing research : JSLHR*, 50(1):119–30, February 2007.

Media Lab Responsive Environments Group. Tidmarsh living observatory. URL `http://tidmarsh.media.mit.edu/`.

Régis Guieu et al. Analgesic effects of vibration and transcutaneous electrical nerve stimulation applied separately and simultaneously to patients with chronic pain. *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques*, 18(2):113–119, 1991.

Eric Gunther. *Skinscape : A Tool for Composition in the Tactile Modality by*. PhD thesis, MIT, 2001.

Stephen Handel. *Listening: An introduction to the perception of auditory events*. The MIT Press, 1993.

P Hansson et al. Influence of naloxone on relief of acute oro-facial pain by transcutaneous electrical nerve stimulation (tens) or vibration. *Pain*, 24(3):323–329, 1986.

Jan Bartošek Václav Hanžl. Comparing pitch detection algorithms for voice applications.

John MacDonald Harry McGurk. Hearing Lips and Seeing voices. *1976 Nature Publishing Group*, 1976.

Elliott Hedman. *Observational Psychophysiology for Empathic Design*. PhD thesis, Massachusetts Institute of Technology, 2014.

Elliott Hedman, Oliver Wilder-Smith, Matthew S Goodwin, Ming-Zher Poh, Rich Fletcher, and Rosalind Picard. icalm: Measuring electrodermal activity in almost any setting. In *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, pages 1–2. IEEE, 2009.

Peter G Hepper, D Scott, and Sara Shahidullah. Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, 11(3):147–153, 1993.

Javier Hernandez, Mohammed Ehsan Hoque, Will Drevo, and Rosalind W Picard. Mood meter: counting smiles in the wild. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 301–310. ACM, 2012.

Dick Higgins. U B U W E B :: Dick Higgins on Sound Poetry U B U W E B :: Dick Higgins on Sound Poetry. pages 1–16, 2010.

Mohammed E Hoque, Joseph K Lane, Rana El Kaliouby, Matthew Goodwin, and Rosalind W Picard. Exploring speech therapy games with children on the autism spectrum. 2009.

Yoshiyuki Horii. An accelerometric measure as a physical correlate of perceived hypernasality in speech. *Journal of speech and hearing research*, 26(3):476, 1983.

John F Houde, Srikantan S Nagarajan, Kensuke Sekihara, and Michael M Merzenich. Modulation of the auditory cortex during speech: an meg study. *Journal of cognitive neuroscience*, 14(8):1125–1138, 2002.

Yeou-Fang Hsieh and Charles H Turner. Effects of loading frequency on mechanically induced bone formation. *Journal of Bone and Mineral Research*, 16(5):918–924, 2001.

William James. *The principles of psychology*. Digireads. com Publishing, 2011.

Tristan Jehan and Bernd Schoner. An audio-driven perceptually meaningful timbre synthesizer. *Analysis*, 2(3):4, 2002.

Elena Jessop. *The body parametric:Abstraction of Vocal and Physical Expression in Performance*. PhD thesis, Massachusetts Institute of Technology, 2014.

FL Jones. Poor breath sounds with good voice sounds. a sign of bronchial stenosis. *CHEST Journal*, 93(2):312–313, 1988.

Camil Jreige, Rupal Patel, and H Timothy Bunnell. Vocalid: personalizing text-to-speech synthesis for individuals with severe speech impairment. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, pages 259–260. ACM, 2009.

Carl Gustav Jung. The Analytical Psychology of. pages 1–13, 1953.

Ellen N Junn, K Morton, and Ivy Yee. The "gibberish" exercise: Facilitating empathetic multicultural awareness. *Journal of Instructional Psychology*, 22(4):324–329, 1995.

Jon Kabat-Zinn, Leslie Lipworth, and Robert Burney. The clinical use of mindfulness meditation for the self-regulation of chronic pain. *Journal of behavioral medicine*, 8(2):163–190, 1985.

Bangalore G Kalyani, Ganesan Venkatasubramanian, Rashmi Arasappa, Naren P Rao, Sunil V Kalmady, Rishikesh V Behere, Hariprasad Rao, Mandapati K Vasudev, and Bangalore N Gangadhar. Neurohemodynamic correlates of 'om'chanting: a pilot functional magnetic resonance imaging study. *International journal of yoga*, 4(1):3, 2011.

Jae-Hoon Kang and Arthur W Leissa. Three-dimensional vibrations of thick spherical shell segments with variable thickness. *International journal of Solids and Structures*, 37(35):4811–4823, 2000.

Sidney Littlefield Kasfir. *West African masks and cultural systems*, volume 126. Musée royal de l'Afrique centrale, 1988.

K Kerschan-Schindl, S Grampp, C Henk, H Resch, E Preisinger, V Fialka-Moser, and H Imhof. Whole-body vibration exercise leads to alterations in muscle blood volume. *Clinical physiology*, 21(3):377–382, 2001.

Adam P Kestian and Tamara Smyth. Real-Time Estimation of the Vocal Tract Shape for Musical Control. 2010.

Youngmoo Edmund Kim. *Singing Voice Analysis / Synthesis*. PhD thesis, MIT, 2003.

Tatsuya Kitamura. Measurement of vibration velocity pattern of facial surface during phonation using scanning vibrometer. *Acoustical Science and Technology*, 33(2):126–128, 2012.

Rebecca Kleinberger. Vocal entropy project webpage. URL `http://fab.cba.mit.edu/classes/862.13/students/rebecca/`.

Jacqueline Kory and Rebecca Kleinberger. Social agent or machine? the framing of a robot affects people's interactions and expressivity.

Sanjay Kumar, HR Nagendra, NK Manjunath, KV Naveen, and Shirley Telles. Meditation on om: Relevance from ancient texts and contemporary science. *International journal of yoga*, 3(1):2, 2010.

Peter Ladefoged. A Figure of Speech: A Festschrift for John Laver edited by William J. Hardcastle, Janet Mackenzie Beck. pages 1–14, 1992.

René Théophile Hyacinthe Laennec. *De l'auscultation médiate: ou, Traité du diagnostic des maladies des poumons et du coeur; fondé principalement sur ce nouveau moyen d'exploration*, volume 2. Culture et civilisation, 1819.

Harlan Lane and Bernard Tranel. The lombard sign and the role of hearing in speech. *Journal of Speech, Language, and Hearing Research*, 14(4): 677–709, 1971.

Marianne Latinus and Pascal Belin. Human voice perception. *Current biology : CB*, 21(4):R143–5, February 2011.

S Levänen and D Hamdorf. Feeling vibrations: enhanced tactile sensitivity in congenitally deaf humans. *Neuroscience letters*, 301(1):75–7, March 2001.

Golan Levin and Zachary Lieberman. In-situ speech visualization in real-time interactive installation and performance. In *NPAR*, volume 4, pages 7–14, 2004.

RP Lippmann. Detecting nasalization using a low-cost miniature accelerometer. *J Speech Hear Res*, 24(3):314–317, 1981.

M. a. Little, P. E. McSharry, I. M. Moroz, and S. J. Roberts. Testing the assumptions of linear prediction analysis in normal vowels. *The Journal of the Acoustical Society of America*, 119(1):549, 2006.

Max A Little. *Biomechanically Informed Nonlinear Speech Signal Processing*. PhD thesis, University of Oxford, 2006.

Jie Liu et al. Biosynthetic response of cultured articular chondrocytes to mechanical vibration. *Research in experimental medicine*, 200(3):183–193, 2001.

TC Lundeberg. Vibratory stimulation for the alleviation of chronic pain. *Acta physiologica Scandinavica. Supplementum*, 523:1–51, 1982.

Raymond MacDonald, Gunter Kreutz, and Laura Mitchell. *Music, health, and wellbeing*. Oxford University Press, 2012.

Tod Machover. Brain opera: Sensor chair, a. URL `http://web.media.mit.edu/~joep/SpectrumWeb/captions/Chair.html`.

Tod Machover. The other festival, media lab arts and design festival, b. URL `http://www.media.mit.edu/events/other/projects`.

Tod Machover. Future arts: By and for whom, and for what? *IBM systems journal*, 39(3.4):678–679, 2000.

Robin Maconie. Stockhausen at 70 Through the looking glass. *The Musical Times, Vol. 139, No. 1863 (Summer, 1998), pp. 4-11*, 139(1863): 4–11, 2013.

Mikhail Malt and Emmanuel Jourdan. Zsa. descriptors: a library for real-time descriptors analysis. *Sound and Music Computing, Berlin, Germany*, 2008.

Philippe Manoury, Omer Corlaix, and Jean-Guillaume Lebrun. *La musique du temps réel: entretiens avec Omer Corlaix et Jean-Guillaume Lebrun.* Editions MF, 2012.

Mika H Martikainen, Ken-ichi Kaneko, and Riitta Hari. Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex*, 15(3):299–302, 2005.

Malia F Mason, Michael I Norton, John D Van Horn, Daniel M Wegner, Scott T Grafton, and C Neil Macrae. Wandering minds: the default network and stimulus-independent thought. *Science*, 315(5810):393–395, 2007.

Peter K. McGregor. Bioacoustics journal. URL `http://www.bioacoustics.info/`.

Jon Mckenzie. Anderson for Dummies. 41(2):30–50, 2009.

David Joel Metzer. The Paths from and to Abstraction in Stockhausen's Gesang der Junglinge. *Modernism/modernity*, 11(4):695–721, 2004.

Marvin Minsky. *Music, mind, and meaning.* Springer, 1982.

Marvin Minsky. *Society of mind.* Simon and Schuster, 1988.

Ashley Montagu. Touching: The human significance of the skin. 1971.

Jerry Moon. The influence of nasal patency on accelerometric transduction of nasal bone vibration. *The Cleft Palate-Craniofacial Journal*, 27(3):266–274, 1990.

Brian CJ Moore and Brian C Moore. *An introduction to the psychology of hearing*, volume 5. Academic press San Diego, 2003.

B Iopac Mpr, Acq Knowledge, Robert Jones, and Mickey Rowe. A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. pages 1–42, 2013.

Nadia Müller, Sabine Leske, Thomas Hartmann, Szabolcs Szebényi, and Nathan Weisz. Listen to Yourself: The Medial Prefrontal Cortex Modulates Auditory Alpha Power During Speech Preparation. *Cerebral cortex (New York, N.Y. : 1991)*, June 2014.

Viktor Müller and Ulman Lindenberger. Cardiac and respiratory patterns synchronize between persons during choir singing. *PloS one*, 6(9):e24893, January 2011.

Raymond LH Murphy Jr, Stephen K Holford, and William C Knowler. Visual lung-sound characterization by time-expanded wave-form analysis. *New England Journal of Medicine*, 296(17):968–971, 1977.

Siegfried Nadel. The origins of music. *The Musical Quarterly*, 16(4): 531–546, 1930.

Nicholas Negroponte. *Soft architecture machines.* MIT press Cambridge, MA., 1975.

Ketty Nez and Hans Tutschku. An Interview with Hans Tutschku. *School of Music 2079 Voxman Music Building University of Iowa Iowa*, pages 14–26, 2003.

Victoria O'Donnell and Garth S Jowett. *Propaganda and persuasion.* Sage Publications, 1986.

Jieun Oh. The Duality of Mechanistic Precision and Timeless Freedom explained in terms of Reaching Identity Goals. pages 1–13, 2009.

George A Ojemann. Individual variability in cortical localization of language. *Journal of neurosurgery*, 50(2):164–169, 1979.

William David Oliver. The Singing Tree: A Novel Interactive Musical Experience. (1995), 1997.

Pauline Oliveros. *Deep listening: a composer's sound practice.* iUniverse, 2005.

World Health Organization. World health organization. URL `http://www.who.int/substance_abuse/research_tools/whoqolbref/en/`.

François Pachet. The Future of Content is in Ourselves. *Csl, Sony Messier, Jean-marie*, pages 1–20, 2010.

George Papcun, Jody Kreiman, and Anthony Davis. Long-term memory for unfamiliar voices. *The Journal of the Acoustical Society of America*, 85(2):913–925, 1989.

Seymour Papert. Teaching children thinking. *Programmed Learning and Educational Technology*, 9(5):245–255, 1972.

J Paradiso, J Smith, M Reynolds, M Orth, R Ciliberto, J Rosenberg, M Bach, B Denckla, P Pelletier, and Knaian. Brain opera: Rhythm tree. URL `http://www.media.mit.edu/~joep/TTT.BO/tree.html`.

Joseph a. Paradiso. The Brain Opera Technology: New Instruments and Gestural Sensors for Musical Interaction and Performance. *Journal of New Music Research*, 28(2):130–149, June 1999.

Rupal Patel, Caroline Niziolek, Kevin Reilly, and Frank H Guenther. Prosodic Adaptations to Pitch Perturbation. 54(August):1051–1060, 2011.

Philip M Peek. the sounds of silence: cross-world communication and the auditory arts in african societies. *American Ethnologist*, 21(3):474–494, 1994.

Corinna Peifer. Psychophysiological correlates of flow-experience. In *Advances in flow research*, pages 139–164. Springer, 2012.

Marcela Perrone-Bertolotti, Jan Kujala, Juan R Vidal, Carlos M Hamame, Tomas Ossandon, Olivier Bertrand, Lorella Minotti, Philippe Kahane, Karim Jerbi, and Jean-Philippe Lachaux. How silent is silent reading? intracerebral evidence for top-down activation of temporal voice areas during reading. *The Journal of Neuroscience*, 32(49):17554–17562, 2012.

B Pfleiderer. Visualization of Auditory Habituation by fMRI. *NeuroImage*, 17(4):1705–1710, December 2002.

Mary Pietrowicz and Karrie Karahalios. Sonic Shapes: Visualizing vocal expression. *ICAD*, pages 157–164, 2013.

Siméon-Denis Poisson. *Mémoire sur l'équilibre et mouvement des corps élastiques*. L'Académie des sciences, 1828.

David W Purcell and Kevin G Munhall. Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, 119(4):2288–2297, 2006.

Y Qi and R E Hillman. Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. *The Journal of the Acoustical Society of America*, 102(1):537–43, July 1997.

LR Rabiner and RW Schafer. Digital speech processing. *The Froehlich/Kent Encyclopedia of Telecommunications*, 6:237–258, 2011.

Herman Rapaport. " Can You Say Hello ?" States Laurie Anderson ' s United. 38(3):339–354, 2013.

Margaret Redenbaugh. Correspondence between an accelerometric nasal/voice amplitude ratio and listeners' direct magnitude estimations of hypernasality. *Journal of speech and hearing research*, 28(2):273, 1985.

C M Reed, W M Rabinowitz, N I Durlach, L D Braida, S Conway-Fithian, and M C Schultz. Research on the Tadoma method of speech communication. *The Journal of the Acoustical Society of America*, 77(1): 247–57, January 1985.

Tobias Riede and Tecumseh Fitch. Vocal tract length and acoustics of vocalization in the domestic dog (canis familiaris). *Journal of Experimental Biology*, 202(20):2859–2867, 1999.

Klaus Biesenbach Rowana Marocci. 2008 Moma Exhibition of Olafur Eliasson work Take your Time, 2001.

Lee Salk. The role of the heartbeat in the relations between mother and infant. *Scientific American*, 1973.

Ladan Shams, Yukiyasu Kamitani, and Shinsuke Shimojo. Visual illusion induced by sound. *Brain research. Cognitive brain research*, 14(1):147–52, June 2002.

Sydney Shoemaker. *Identity, cause, and mind: Philosophical essays*. Oxford University Press, 2003.

Rahul Shrivastav and Christine M. Sapienza. Objective measures of breathy voice quality obtained using an auditory model. *The Journal of the Acoustical Society of America*, 114(4):2217, 2003.

Yen-Liang Shue and Abeer Alwan. A new voice source model based on high-speed imaging and its application to voice source estimation. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 5134–5137. IEEE, 2010.

D G Sinex and C D Geisler. Auditory-nerve fiber responses to frequency-modulated tones. *Hearing research*, 4(2):127–48, May 1981.

John Smalley. Gesang der Jünglinge : History and Analysis. pages 1–13, 2000.

Viola Spolin. *Improvisation for the theater: A handbook of teaching and directing techniques*. Northwestern University Press, 1999.

Alex Stahl and Patricia Clemens. Auditory Masquing : Wearable Sound Systems for Diegetic Character Voices. (Nime):15–18, 2010.

Kenneth N Stevens. *Acoustic phonetics*, volume 30. MIT press, 2000.

Julian M Stewart et al. Plantar vibration improves leg fluid flow in perimenopausal women. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 288(3):R623–R629, 2005.

Johan Sundberg. *Chest vibrations in singers*, volume 22. Centre Georges Pompidou, 1979.

Hong Zhang Tan, Nathaniel I Durlach, William M Rabinowitz, and Charlotte M Reed. Information transmission with a multi-finger tactual display. *Scandinavian audiology. Supplementum*, 47:24–28, 1996.

Nick Tsakalos and Evangelos Zigouris. Autocorrelation-based pitch determination algorithms for realtime vocoders with the TMS32020/C25. *Microprocessors and Microsystems*, 14(8):511–516, October 1990.

Hitomi Tsujita and Jun Rekimoto. Smiling makes us happier: enhancing positive mood and communication with smile-encouraging digital appliances. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 1–10. ACM, 2011.

James Turrell. Aten Reign (2013),, 2013.

Kleinberger Rebecca Van Troyer Akito. Miniduck webpage. URL `http://web.media.mit.edu/~akito/Projects/MiniDuck/`.

Ronald T Verrillo. Vibration sensation in humans. *Music Perception*, pages 281–302, 1992.

Björn Vickhoff. A perspective theory of music perception and emotion. *rapport nr.: Skrifter från musikvetenskap 90*, 2008.

Björn Vickhoff, Helge Malmgren, Rickard Å ström, Gunnar Nyberg, Seth-Reino Ekström, Mathias Engwall, Johan Snygg, Michael Nilsson, and Rebecka Jörnsten. Music structure determines heart rate variability of singers. *Frontiers in Psychology*, 4(July):1–16, 2013.

Katharina von Kriegstein, Evelyn Eger, Andreas Kleinschmidt, and Anne Lise Giraud. Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain research. Cognitive brain research*, 17(1):48–55, June 2003.

Ludwig Wittgenstein. The blue and brown books. 1958.

David B Wolf and Neil Abell. Examining the effects of meditation techniques on psychosocial functioning. *Research on Social Work Practice*, 13(1):27–42, 2003.

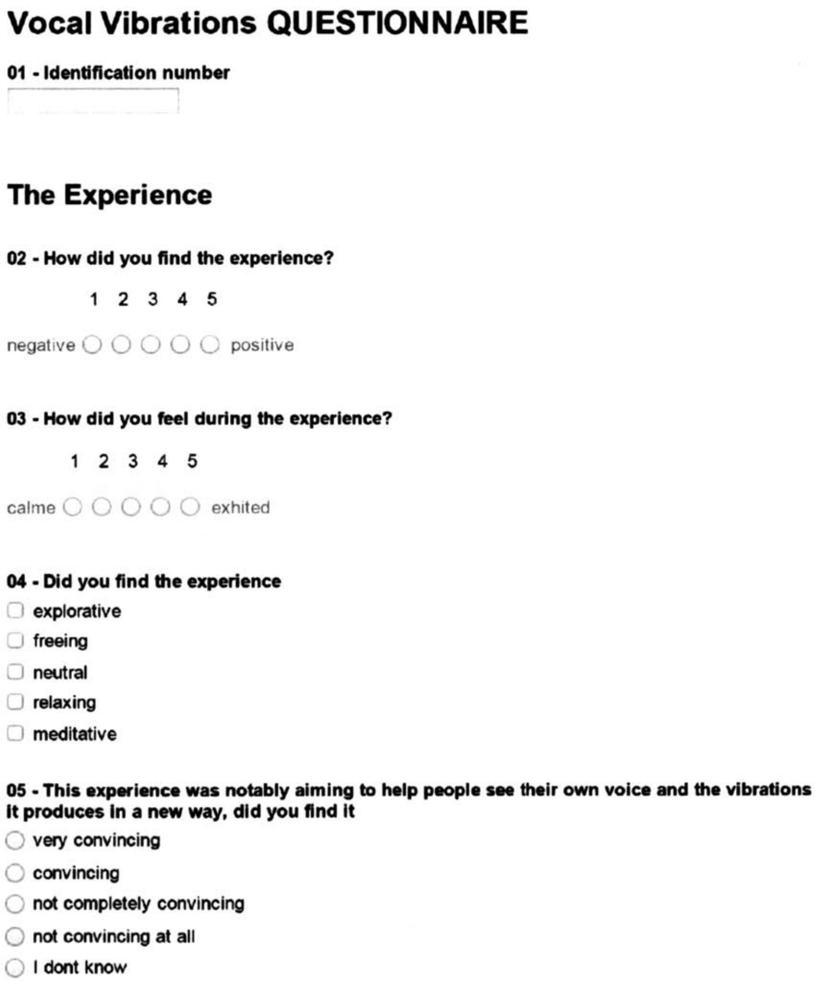
Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics (TOG)*, 31(4):65, 2012.

Matías Zanartu, Julio C Ho, Steve S Kraman, Hans Pasterkamp, Jessica E Huber, and George R Wodicka. Air-borne and tissue-borne sensitivities of bioacoustic sensors used on the skin surface. *Biomedical Engineering, IEEE Transactions on*, 56(2):443–451, 2009.

Zhihong Zeng, Maja Pantic, Glenn I Roisman, and Thomas S Huang. A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, 31(1):39–58, January 2009.

Appendix

Questionnaire used in Measurement of Vocal Practices



Vocal Vibrations QUESTIONNAIRE

01 - Identification number

The Experience

02 - How did you find the experience?

1 2 3 4 5

negative ○ ○ ○ ○ ○ positive

03 - How did you feel during the experience?

1 2 3 4 5

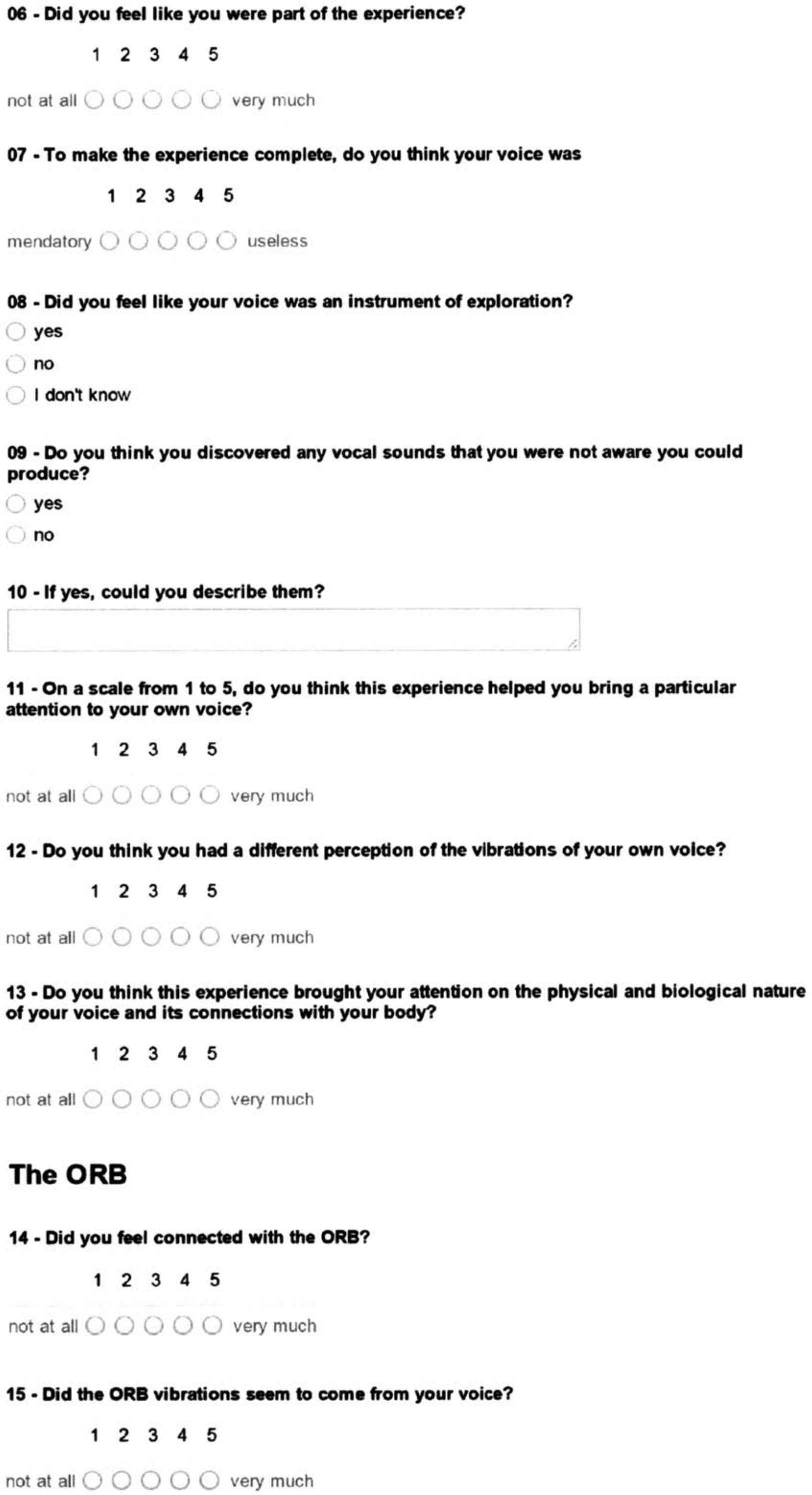
calme ○ ○ ○ ○ ○ exhited

04 - Did you find the experience

- ☐ explorative
- ☐ freeing
- ☐ neutral
- ☐ relaxing
- ☐ meditative

05 - This experience was notably aiming to help people see their own voice and the vibrations it produces in a new way, did you find it

- ○ very convincing
- ○ convincing
- ○ not completely convincing
- ○ not convincing at all
- ○ I dont know



06 - Did you feel like you were part of the experience?

1 2 3 4 5

not at all ○ ○ ○ ○ ○ very much

07 - To make the experience complete, do you think your voice was

1 2 3 4 5

mendatory ○ ○ ○ ○ ○ useless

08 - Did you feel like your voice was an instrument of exploration?

○ yes
○ no
○ I don't know

09 - Do you think you discovered any vocal sounds that you were not aware you could produce?

○ yes
○ no

10 - If yes, could you describe them?

11 - On a scale from 1 to 5, do you think this experience helped you bring a particular attention to your own voice?

1 2 3 4 5

not at all ○ ○ ○ ○ ○ very much

12 - Do you think you had a different perception of the vibrations of your own voice?

1 2 3 4 5

not at all ○ ○ ○ ○ ○ very much

13 - Do you think this experience brought your attention on the physical and biological nature of your voice and its connections with your body?

1 2 3 4 5

not at all ○ ○ ○ ○ ○ very much

The ORB

14 - Did you feel connected with the ORB?

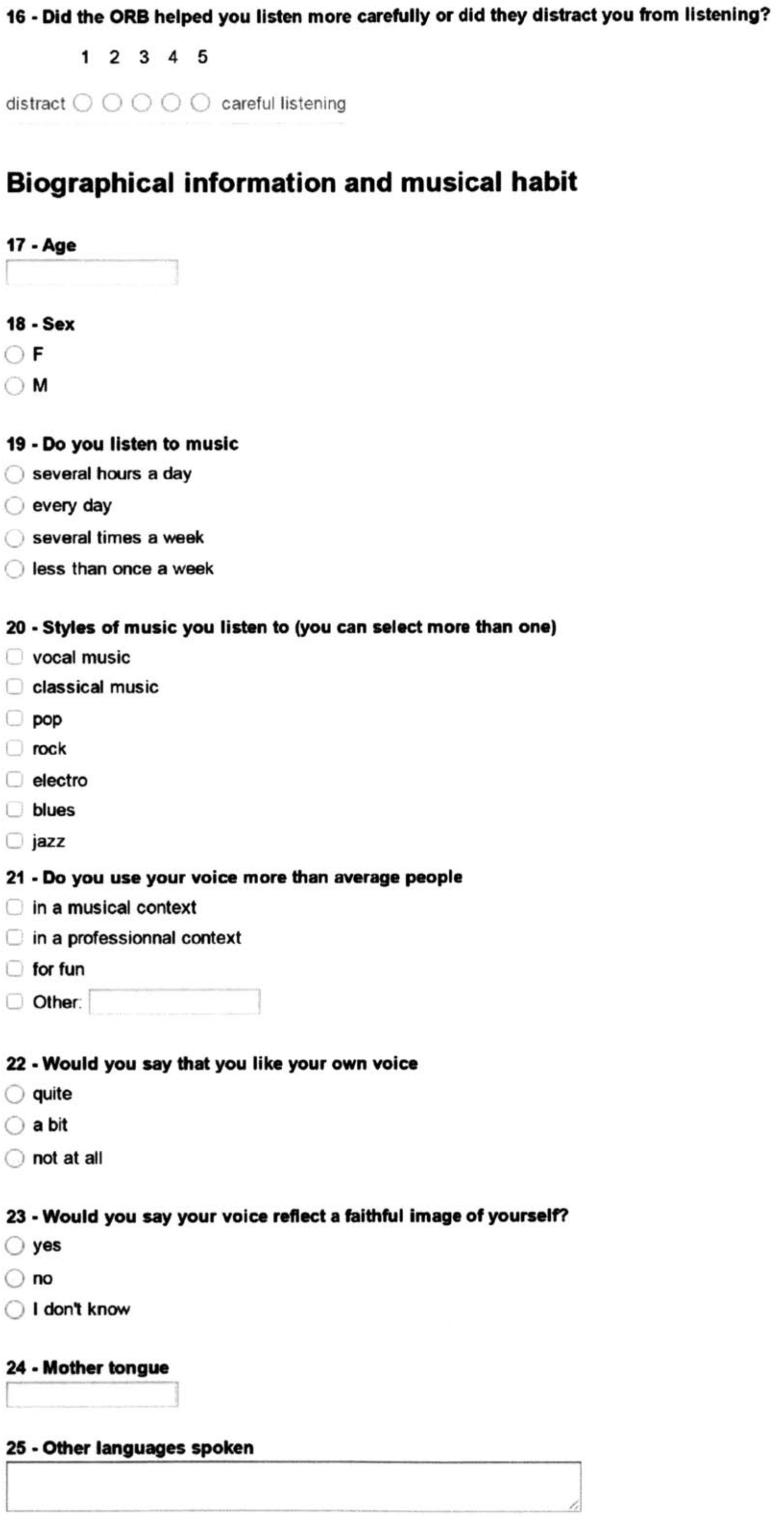
1 2 3 4 5

not at all ○ ○ ○ ○ ○ very much

15 - Did the ORB vibrations seem to come from your voice?

1 2 3 4 5

not at all ○ ○ ○ ○ ○ very much



16 - Did the ORB helped you listen more carefully or did they distract you from listening?

1 2 3 4 5

distract ○ ○ ○ ○ ○ careful listening

Biographical information and musical habit

17 - Age

18 - Sex

○ F
○ M

19 - Do you listen to music

○ several hours a day
○ every day
○ several times a week
○ less than once a week

20 - Styles of music you listen to (you can select more than one)

☐ vocal music
☐ classical music
☐ pop
☐ rock
☐ electro
☐ blues
☐ jazz

21 - Do you use your voice more than average people

☐ in a musical context
☐ in a professionnal context
☐ for fun
☐ Other:

22 - Would you say that you like your own voice

○ quite
○ a bit
○ not at all

23 - Would you say your voice reflect a faithful image of yourself?

○ yes
○ no
○ I don't know

24 - Mother tongue

25 - Other languages spoken