ABSTRACT

Vocal Vibrations is a new project by XXX’s research group that 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 paper describes the motivations, the technical implementation, and the experience design of the Vocal Vibrations public installation. This installation will consist 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 musical environment is shaped live by the participant’s vocal gestures. Simultaneously, the participant 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 context. Through these vocal explorations, they can discover new aspects of their voices and gain a deeper appreciation of and for their everyday instrument.

Keywords
Voice, Vibrations, Expressive Interfaces, XXX, Signal Processing, Public Installations, Tactile Interfaces ...

1. INTRODUCTION: VOCAL VIBRATIONS

The experience that a person has of his or her voice is quite intimate. It is infinitely expressive and individually defining. However, many people today do not generally explore their vocal abilities, do not typically pay close attention to their voices, and do not feel comfortable “singing” or imagine they could participate in a rich musical experience through their voice. Our brain even responds less to our own voice than it does to other voices [16]. In the Vocal Vibrations project, we aim to guide participants in exploring a wide range of vocal sounds and vibrations. To address this, we are developing techniques to engage the public in the regular practice of thoughtful singing and vocalizing, both as an individual experience and as part of a community. The first experience resulting from this work will be a public interactive installation at a collaborative design studio. We are designing the installation to provide two contrasting spaces: one for a group experience, one for a private experience. Visitors will arrive in a larger room we call the Chapel, a venue for listening closely to a recorded electroacoustic composition centered on the singing voice. The inner room, the Cocoon, provides a private isolated environment, inviting visitors to sing and participate in an interactive vocal experience. This initial Vocal Vibrations installation aims to raise awareness of the influence of the voice on our body and environment, as well as to give participants the ability to experience their voices in a new light, by enabling anyone to control a rich multi-sensory experience with only his or her voice.

One of the goals of this project is to help even novices discover the potential of their voice while providing them with direct access to the richness of a full musical experience. Indeed, with appropriate analysis and interactive systems, a user can become part of a complex vocal performance and have an active role in shaping the musical result through even the simplest use of his voice, such as exploring variations on a single pitch. While a person vocalizes alone in the Cocoon, the experience offers two components. First, the musical environment reacts to and accompanies the user’s vocal gestures. Second, people experience an exteriorization of their voice through a handheld device, the Oral Resonant Ball (ORB), that maps their voice into the tactile sensation of vibration. For the interactive auditory and tactile experience, we are designing the system such that there is no “right” or “wrong” way to use it. The system built for this project processes the user’s voice signal in real time to extract features that are transposed into control parameters for an interactive experience. The system accompanies the user’s vocal explorations, while also guiding the user to extend these explorations. We seek to expand our group’s work in technologies for sophisticated measurement and extension of the singing voice in performance to create new kinds of vocal experiences in which everybody can participate.

We also seek to bring attention to the nature of the voice as an incredibly physical instrument. The act of singing and vocalizing creates vibrations throughout the body that can be altered through modifications to vocal production. 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 [25]. The focus provided by tactile and physical feedback can help to give intimate,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
while still objective, access to the voice. As the ORB externalizes the vibration of a user’s voice, it thus can help the user become more aware of the variation and range of that vibration. In exploring the relationships between human physiology and the resonant vibrations of the voice, we seek to address many questions related to the voice and its connection with the body as well as its influence on mental and physical health.

2. RELATED RESEARCH ON THE VOICE

2.1 Voice, Body, Mind, and Vibration

Most of us do not pay attention to the complex physical processes involved in producing a vocal signal, particularly one that is expressive or emotional. Additionally, the use of our voice is a goal-directed activity. All the complex psychomotor sub-processes are activated without conscious separation [23]. Yet, neurological research supports the idea that the brain dissociates voice from speech when processing vocal information [45]. 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 [16]. 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.

Because the voice requires a perfect psychomotor synchronization between many physical processes (such as the breath, the tongue, the vocal tract muscles, the tension of the vocal folds, and the lips), the study of the voice can reveal details about a person’s health and mental state [1]. Mental and emotional states are often apparent through the voice because the physical procedure is closely shaped by emotions. Kenneth Stevens [40] describes those correlates in terms of vocal modification in situations of strong arousal. For example, in the case of stress, variations in muscle contraction and breathing patterns have a direct influence on the sound of the voice. Max Little implemented a biologically and mechanically based model that takes into account nonlinearity, non-Gaussian and turbulent aspects present in vocal production [31]. 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 [30].

Not only can studying the voice reveal information about physical, mental and emotional states, but using the voice can also affect those states. 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) [8, 14, 33]. Those studies generally agree on the fundamental importance of breath control, induced by the use of the voice, as an important connection between singing and physiology.

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 [18, 9]. 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 [19].

2.2 Measuring Vocal Emotion and Expression

A key aspect of our research is how to recognize the affective, expressive, and personal content of an individual’s voice, and to determine what features of the voice help convey that information. Particularly in the domains of speech recognition and detection of vocal dysfunction, studies have explored descriptive frameworks for vocal qualities [26, 39]. Scherer identifies a variety of vocal features that convey expression, including vocal perturbation (short-term variation), voice quality (timbre), intensity, tempo, and the range of the fundamental frequency over time. Other research has focused on affective markers present in the voice. Cahn’s work on generating expression in a synthesized voice [6] offers a good overview of the acoustic and prosodic aspects that correlate to emotions, including basic voice parameters that are perceptually important to convey expressivity (pitch, timing, voice quality and articulation). Fernández explores the mathematical description of voice quality, highlighting the deconvolution of the speech signal into the glottal flow derivative and the vocal tract filter to be able to mathematically access the emotional quality of speech [11].

Other researchers have done analyses specific to the singing voice with the goal of developing better algorithms for singing voice synthesis [21] or using the singing voice as input to another synthesis algorithm [42, 17, 12]. Kim separates features reflecting an individual’s vocal physiology (such as the configuration of the vocal folds and vocal tract) from features reflecting an individual’s expressive performance (such as how those features vary over time). Features of the singing voice have also been used for synthesis models, such as Chréode, a computer generated tape piece using the CHANT program from IRCAM. Conceived as an interactive instrument, the CHANT synthesizer is based on physics but controlled by perceptual parameters, such as frequency of the fundamental, random variations of the fundamental, vibrato, random variations of the vibrato, spectrum, formants and fundamental, etc [38].

However, when it comes to extracting meanings or features from the voice, only a few feature extraction tools are adapted to the specific range, richness and complexity of the human voice. Our work is drawn from some research on extracting prosody features [24, 44], vocal quality elements [20, 21, 35], and affective markers [2, 6, 11] from the vocal signal.

2.3 Prior Expressive Vocal and Vibrational Experiences

Prior interfaces for manipulating the voice in performance include handheld devices like Waitsvisa’s The Hands [4]; systems for changing vocal timbre such as those used by Laurie Anderson [36]; and wearable systems such as Sonami’s Lady’s Glove [5], the Bodycoder system [3], and the BodySynth used by Pamela Z. [29]. However, the majority of these vocal manipulations are controlled by external buttons, key-boards, or a performer’s movement, rather than solely by the parameters of the vocal input as occurs in the Vocal Vibrations installation.

Levin and Lieberman have also incorporated graphics shaped by vocal production into public installations in Messa di Voce, Hidden Worlds, and RE:MARK [28] (In these experiences, the amplitude and spectral content of visitors’ voices were used to affect projected graphics. Another public installation focusing on the voice is Oliver’s Singing Tree [34], with which visitors interacted through singing into a microphone. The “pitch, noisiness, brightness, volume, and formant frequencies” of their voices were measured, and these parameters were used in real time to control a music gener-
ation engine and a video generation system. All of these installations have a strong playful component, with the goal of an interesting vocal experience. Focusing carefully on subtle variations of sound has been a component of new music compositions, such as Lucier’s “I Am Sitting in a Room” and Chowning’s “Phon´e” [32, 7] but not of an interactive vocal installation.

Prior work also explores the possibilities of rich aesthetic experiences centered around vibration. Skinscape [15] is a tool for composition in the tactile modality. 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. Our work to create vibration experiences derived from and driven by the voice is inspired by this research.

3. THE VOCAL VIBRATIONS INSTALLATION

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. When a participant first arrives at the Vocal Vibrations installation, she will enter a communal space designed for close listening. In this space, which we call the Chapel, the audio will be a precomposed electroacoustic composition by XXX based on recordings of voices. Each participant will then be approached individually by an assistant who will bring him or her to a smaller space in preparation for the interactive vocal experience. This assistant will instruct the participant to vocalize on one pitch, and will help the participant to find a note that fits comfortably in her range. This step will help the user get used to the note and may also “tune” the system to the person’s voice and specific note. The participant will then be brought to a structure specially designed by YYY, the Cocoon, where she will be invited to sit, given headphones and the vibrating ORB to hold, and left alone in the space. The participant will then have a solo experience, approximately five minutes long, where both the sound played in the headphones and the behavior of the ORB will be controlled and shaped by her vocal explorations. The audio will be an interactive piece inspired by the composition in the Chapel, with a fixed structure but with the flow and variations of the composition controlled by the user. At the end of the solo experience, the user will return to the Chapel, where she is free to stay and listen as long as she wishes, as well as to vocally improvise along with the music if she desires. All of the musical content in this installation is new material composed by XXX.

3.1 The Chapel: Focused Listening

When visitors first arrive at Vocal Vibrations, they will enter the outer chamber, the Chapel, intended for a quiet, meditative experience. Here, in a new composition by XXX, singing voices will surround visitors and gently envelop them in sound. Visitors can remain in this space for as long or short as they desire, choosing to join in through humming or vocalizing or simply to listen. The composition in the Chapel will be assembled from many layers of pre-recorded solo and choral vocal material, designed such that a D is almost always present in the score. Speakers will be located around the room such that the composition can be spatialized.

An important part of this project, particularly in the Chapel, is its use of surround sound techniques to put a visitor in the midst of the musical experience. Conventional sound reinforcement systems use loudspeakers to convey an electronic audio signal as clearly as possible. For Vocal Vibrations, we developed a conceptual model and software implementation that couples loudspeaker configuration with digital instrument design. Each loudspeaker is treated as a signal processing node in 3D space that is connected to its nearest neighbors. Each node has an identical set of sound processing capabilities, the ability to send instructions to neighboring nodes on how to apply these processing capabilities, and the ability to follow instructions received from neighboring nodes.

For example, a speaker node might receive and follow these instructions:

- Begin playing AudioSample.wav
- Ramp volume of AudioSample.wav from 0. to 1. over 300 milliseconds
- Wait 300 milliseconds
- Broadcast this instruction to all neighboring speaker nodes that have not yet received this instruction

By following instructions like the one described above, we create surround sound experiences that swirl and envelop.

3.2 The Cocoon: Interactive Vocal Experience

In the second portion of the installation, a private environment, the Cocoon, will allow individual visitors to have a meditative experience exploring the vibrations generated by their own voice, accentuated within this space through acoustic and physical stimuli. The technical systems used for this installation include: a real-time system for processing vocal signals, developed in Java; a flexible Java-based mapping system that determines output control parameters from low-level extracted features and high-level parameters describing voice quality and vocal gesture; a Max/MSP patch that controls the behavior of the sound system (including composition choices, samples, localisation, effects); and a Max/MSP patch that controls the vibration behaviors of the ORB. All systems communicate via the Open Sound Control protocol.

3.2.1 A Personal Experience

From the Chapel, each participant will be guided by an assistant into the interactive experience. A short “training” session will follow, in which the participant is encouraged
to take the first step into producing vocal sounds. The assist-
tant first will assess the frequency range of the participant’s
voice and then give him a D in the most comfortable octave.
The participant will be asked to hold the D and will be given
the 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 ex-
periment with a wide range of vocalizations, including vari-
atations on extended vocal techniques (such as Sprechgesang,
inhalting, tremolo, overtones, and changing the shape of the
mouth) as well as even non- or semi-voiced sounds (such as
breathing and whispering).

From this simple entry point, vocalizing on a single note,
a participant can take control of an interactive musical piece
based on the longer composition in the Chapel, as well as a
corresponding vibration experience in a handheld device,
the ORB. At different moments, short sentences will ap-
pear on a screen in front of the participant to invite him to
vocally explore in a particular way (“Like the surrounding
sound,” “In an unexpected way,” etc.). In this exploration,
there is no “right” or “wrong” way to vocalize, as the inter-
active experience is constrained by composition choices and
mapping decisions.

For the system to be interactive, the first step is to extract
a certain number of meaningful features as well as the raw
signal from the voice. One of the important steps was de-
termining which kinds of parameters we wanted to measure
from the voice. We determined that there are two cate-
gories of relevant information: low-level parameters (such as
frequency, amplitude, and harmonicity) and high-level
parameters that can be abstracted from the voice (such as
energy and complexity).

First, as the participant vocalizes in a microphone, the
raw signal of his or her voice is used in real time as part of
the behaviours of the ORB. The first level of control pa-
rameters we extract are: pitch, loudness, linearly averaged
frequency spectrum, and harmonicity. These are computed
by spectral analysis. Our objective in this choice of pa-
rameters 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.

### 3.2.3 Interactivity and High-Level Parameters

In addition to the vocal analysis parameters listed above, we
also are interested in a variety of abstract, high-level param-
eters describing the “quality” of the voice or vocal gestures,
such as energy, complexity, fluidity, intensity, and rate. To
obtain these abstract parameters, we are incorporating the
Expressive Performance Extension System, a tool designed
for flexible mapping of input data streams to output control
parameters through a node-based visual language [43, 10].
This system has been extended for the analysis of move-
ment and vocal qualities [10]. It allows users to obtain raw
input data, extract expressive features, define desired vocal
and physical qualities, perform pattern recognition to iden-
tify those qualities, and manually map information about
these high-level expressive parameter spaces to output con-
tral parameters of the interactive experience. In all cases,
the outputs of the pattern recognition algorithms are con-
tinuous values, not classification; the goal is not to label a
gesture “staccato” or “fast,” but to find a position on a
set of continuous expressive axes. Using the Expressive
Performance Extension System, we can explore how best
to interpret that data and turn it into expressive informa-
tion that is useful for creating interactive performances and
installations.

In Vocal Vibrations, the low-level vocal parameters are
used as inputs to pattern recognition processes within EPES
that define a participant’s current vocal expressivity in terms
of high-level parameters. A combination of intermediate
and high-level parameters are then used to shape the way
in which the sonic environment responds to the user’s voice,
with the goal of creating audio accompanying that is not
only immediate and clear, but also satisfyingly complex.
The format of the resulting experience will be a blend of a
pre-composed experiential structure with moment-to-moment
variations around that structure shaped by the vocal explo-
iers of the solo participant.

### 3.3 The ORB

As part of this project, we have also built the Oral Reso-
nance Ball (ORB), a voice-activated vibrating device, which
maps a vocal signal into tactile sensations. This device is
designed to provide awareness of the physical processes
involved in vocal production by giving feedback about and
enhancing the vibrations produced in a person’s body. Fin-
gertips contain more sensor receptors than our vocal vibrat-
ing chamber [22, 27]; thus, the same vibrational signal sent
into the hands will be felt differently and with more de-
tail than when sent into the body. We have found that the
hands can detect many variations in vibration caused by
amplitude, frequency, and timbre. Additionally, research
on the Tactaid (a tactile hearing aid that codes sound in-
formation via a set of vibrators resting on the forearm) has
shown that vibration enhances lipreading performance in
hearing impaired individuals [13].

Holding the ORB in one’s hands while vocalizing can give
one access in another medium to detailed elements from the
voice that often remain latent in one’s everyday experience
of voice. Additionally, making the vibration of the voice
something that can be experienced externally is intended to
connect people to their voice in a new way. We offer 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.

#### 3.3.1 Hardware

The ORB is a ovaloid shaped frosted glass shell measur-
ing about 15 by 15 by 10 centimeters, with five transducers
attached on the inside wall. The materials and precise set-
tings are chosen to maximise 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 vibra-
tional effects on the hands and fingers. ([Because glass has

Figure 2: Prototype of The Cocoon Designed by YYY
the microstructural property of presenting no directional atomic order, the material offers the beneficial properties of smoothly blending the vibration from one transducer to another while keeping certain localized effects.))

3.3.2 Behavior

The control system for the ORB consists of a Max/MSP patch that sends a processed signal to each of the 5 localized channels based on a set of control parameters. In addition to the raw vocal signal that is sent to the top channel for intuitive tactile feedback, the ORB also vibrates with additional textures of dynamic tactile patterns on the surface of the shell. The textures are made of granular tactile signals with specific behavior of location, speed and scattering around the surface, creating abstract impressions like purring, hatching, whirling, etc. The real time feature extraction system and the Expressive Performance Extension System described above are used to control the dynamics and localization of the vibrations in the device. The ORB control parameters also shape additional effects on the signal such as delay, attenuation and a light feedback. The mapping of the ORB will change at moments during the interactive experience.

![Figure 3: The ORB: System Diagram](image)

The skin’s response to stimuli is not linear. When coding the behaviors of the Orb, we have had to take into account that the signal sent to the Orb is subjected to three serial, non-linear sources of physical alterations before being perceived: the transducers, the material of the shell, and the skin of the user’s fingers and palm. The nonlinearity of the transducers is resolved by tuning them through applying a different gain to each of the five signals. We also will adjust the strength of the signals in accordance with Stevens’ Power Law [41]. In studies of sensitivity to vibrations and other tactile stimulus, it has been established that tactile sensitivity to vibration is highly dependent on the frequency of that vibration. Additionally, the range of the vibrotactile frequency response to which skin is sensitive is 20 - 1000 Hz. This range is much narrower than the auditory frequency range our ears can detect (20 - 20,000 Hz). Thus, the frequencies of signals sent to the Orb should differ from an audio signal in order to be perceived through touch.

4. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we have described the Vocal Vibrations project, designed to encourage people to explore and pay thoughtful attention to the range of their vocal sounds and vibrations, to have a rich musical experience centered on their voice, and to experience their voices in a new way. We have also described the first public installation being developed for the project, including a space for careful, meditative listening and a space for interactive vocal exploration. This initial Vocal Vibrations installation will premiere at ZZZ in Paris in March 2014, and remain installed for five months.

The Vocal Vibrations project is part of a larger research initiative around the human voice that our group is undertaking. This initial installation will serve as guidance as we develop our future research directions. Through examining the experience of the audience, we will observe to what extent the installation’s flow, technologies and interactions allow the public to feel engaged. The design of the installation will include preparatory stages of testing the system on peers and iteration on our design based on those tests. Our discussions and assessment of both the initial experiments and the final installation will revolve around several aspects: the quality of the overall experience; the reactivity of the system; the feeling of connection to and control by the participant’s voice; and the coherence of the experience. The final development of the March installation will also reveal the weaknesses and strengths of the overall experience. Those observations will guide us as we design a second version of the Vocal Vibrations installation intended to open in Cambridge in the fall of 2014.

We also seek to expand our explorations of the vibrations tied to the voice and methods for transforming a participant’s experience of those vibrations. In this effort, tools built for the deaf community can also be an interesting source of inspiration, such as the Tadoma method of “tactile lip reading” [37], where a deaf person uses their hand to pick up vibrations and movement from the speaker’s lips, jaw, cheek, and throat. This use of alternative senses to get as close as possible to the physical process voice production is inspirational because it also brings people closer to the emotion and liveness of the voice.

Future work tied to Vocal Vibrations will also include a series of multimedia experiences, including more individual “meditations” group singing experiences, and even personal mobile applications, all designed to help participants explore their individual, expressive voice and its effects on the body through an enveloping context of immersive, responsive music.

5. ACKNOWLEDGMENTS

Acknowledgements to be added when paper is no longer anonymized

6. REFERENCES


