Reading Texts in Real Time:
The Dramatic Voice in/of Interactive Media

Joseph ROVAN

email: rovan@music.unt.edu

Assistant Professor and Director, CEMI

Center for Experimental Music & Intermedia

University of North Texas

Denton, Texas 76203

Katherine BERGERON

email: bergeron@socrates.berkeley.edu

Associate Professor

Department of Music

University of California, Berkeley

Berkeley, California 94720

Abstract

What is the cause, and the cost, of insight? What does it mean to see the world face to face? These were questions raised by the young Rilke in a well-known passage from his *Notebooks of Malte Laurids Brigge,* a Parisian diary he kept in 1904 while serving as personal secretary to Rodin. They are also questions lying at the heart of *Vis-à-vis*, a multimedia work for voice, live electronics and real-time video, which takes Rilke's words as a dramatic point of departure.

This paper begins by addressing the formal and psychological aspects of "Vis-à-vis" as a way of understanding the piece's attitude towards, and use of, interactive technology. Performing within a dense sonic landscape whose electronic sounds are based entirely on her voice, the solo singer interacts, in effect, with a computer to convey the drama of Rilke's unusual text. The interactive score, written in MAX/MSP, makes the computer a virtual performer, listening and responding to the changing musical events in its own terms. A level of independence—or "obstinance"—designed into the real-time environment creates an ever-changing system.

Together with the electronic score, the piece's five scenes are illuminated by a corresponding visual score, assembled in real time by means of another custom program,

written in the video-programming environment Onadime. Much of the visual material is, as in the audio score, based on different aspects of the singer's visage. The exception is a single visual quotation, a haunting image of a woman caught unawares by the turn-of-the-century Parisian photographer Eugène Atget. The aim was to deploy the two layers of interactive media to suggest the dramatic structure itself—the shifting psychological relationship between the woman who sings (the "I" narrating the story) and the nameless woman who is ultimately the subject of Rilke's narration. Incorporating the creative perspectives of composer and performer, this paper will approach the question of reading Rilke as both text and act.

Rilke's faces: The Drama in the Text

Drawn from diaries Rilke kept while he was living in Paris in 1904, the little story "Gesichter" ("Faces") reads in many ways like a prose poem. The text amounts to a brief meditation on the nature of seeing, told in five paragraphs and from just about as many points of view. The poet begins unexpectedly with a question, as if addressing himself to an absent interlocutor, or perhaps to no one at all: "Habe ich es schon gesacht?" (Have I said it before?). The question seems to interrupt the flow of a prior contemplation to make way for a more significant confession, for Rilke immediately tells us, "Ich lerne sehen" (I am learning to see). We do not yet know what such learning entails, but we are told that it too is a beginning, and, like all beginnings, it is "still going badly." It is tempting to think that the poet, long accustomed to looking at the world through the art of metaphor, had suddenly embarked on an entirely new visual education when he became the personal secretary to one of France's most prominent sculptors. Indeed, it seems that one thing the writer had never noticed is the depth of expression hidden in the human face.

"It never occurred to me before how many faces there are, " he writes, with more than a touch of irony. And that observation launches a brief, and mockingly philosophical discussion of the malleability of expression. Although people have many faces, he says, one can still find some thrifty types who would prefer to save their inheritance. They

wear the same face year after year, so that it becomes threadbare, stretched out like old leather gloves. Yet even while putting the others away for posterity, the unthinkable can happen. The dogs could get to them first, Rilke concludes. His little joke is obvious, suggesting a curious rationale for the uncanny resemblances we've all noticed between people and their pets. The humor soon turns darker, however, as Rilke contemplates the opposite phenomenon: the wasteful, overly expressive person who changes his face too often. Easy come, easy go, the poet seems to say. "Why, he's barely forty years old before he's gone through the whole supply!" And when that happens, the poor soul will have no choice but to go around—unimaginably—with nothing on at all: wearing, as Rilke calls it, the "non-face" (das Nichtgesicht).

The next moment in the meditation, however, puts a stop to the philosophizing, and to all the irony, as the poet focuses on something far more real. The text makes a sudden shift from the present tense to the past; from humor to horror; fiction to history. The change of perspective seems to come without any preparation at all, introduced by nothing more than a conjunction. "But the woman," he says, as if beginning all over again. "The woman." ("Aber die Frau" . . . "die Frau.") Whom does he mean? We don't yet know. Rilke himself does not actually know her. But it is clear that she has been there all along in his story, as the ostensible source of his visual education. He met her "on the corner of rue Notre-Dame des Champs." A spot of double-blindness, for neither one has seen the other. She was poor. She was lost in thought. His footstep startled her. It was then, however, that he had his own startling moment of insight. The woman looked up suddenly—so suddenly, in fact, that he *sees something he has never seen before*: he glimpses, in effect, the face of the faceless poor, and the vision is monstrous. For the frightened woman, now left with her face in her hands, presents a sight far more alarming than anything the poet has ever experienced. To look her in the eye was to see the horrible, naked truth: a person without any face at all, *ohne Gesicht*.

In five dense paragraphs Rilke has presented us, then, with a tiny drama of seeing, a drama that turns around a haunting—and hauntingly absent—character, one who happens to have been seen only by the poet. This story of perception is also a story about change,

and the difference one fleeting vision can make in a human life. The "I"/eye who has seen is now different, although it still has (we are told) a lot to learn. It was ultimately this uncertain drama of the self—a drama built around this vision of absent presence—that we hoped to capture by collaborating on *Vis-à-vis*, collectively exploring the question, and the burden, of insight through the musical and visual tableaux, through the interactive technology, and ultimately through performance. We shall begin with a few remarks from the composer about translating Rilke's text into music.

The Composer's Voice*: Reading Rilke as Music

The complex logic of Rilke's self-reflexive text naturally became an important framework for the composition, both from a musical and technical standpoint. While I had already composed interactive works based around a text (for example, my "Continuities" for glove controller and interactive electronics, performed at the Connecticut College *Symposium on Arts and Technology* in 1999), I had never before used a text that demanded such a dramatic reading. In "Continuities," a tiny poem by Archie Ammons functioned, so to speak, as a *pretext* for composition, a quantity of musical material culled for its sonic potential rather than its poetry. While essential to the piece at one level, the poem *as* poem was also, in some respects, left behind. In approaching Rilke, however, I wanted above all to leave the prose intact. Because the dramatic logic of the text—with its unexpected twists and sudden changes of affect—was essential to the moment of insight it described, I needed to find a way to make the text audible throughout the composition, to make it comprehensible for the audience. I needed, in short, to develop an approach to the electronics that would not only allow the words to "speak," but would also draw out their dramatic impact. The problem became one of balancing a desire to preserve the clarity of Rilke's drama, with its flash of insight, against technology's inherent tendency to obscure. Achieving this balance was a long process of trial and error, which perhaps explains why the work took nearly two years to complete.

Thus, the text came first. Moreover, the very first words took on a kind of primary importance: the opening, rhetorical question, "Habe ich es schon gesagt? (Have I said it before?), and its equivocal answer, "Ich lerne sehen." (I'm learning to see). This pair of statements, with their suggestion of possibly failed memory and perception, became a driving force behind the composition. The most obvious manifestation occurs at the level of form. The announcement "Ich lerne sehen" recurs as a series of refrains that progress over the course of the drama, to reflect the psychological development of the first person. The voice that, at the end of the piece, reaches out to the audience in a naked cry of pain reveals the depth of this sentimental education. It is, in short, a far cry from the profound and measured resignation of the chant that opens the piece. The enormous distance between the two refrains thus raises a question: who is the wiser self? The "ich" of the beginning or that of the end?

That question ultimately speaks to a deeper level of the composition, for the overall trajectory of the piece can be seen, in a sense, to move forwards and backwards at the same time. Just as the subject's moment of insight depends on an act of reflection that escapes the forward momentum of the narrative retelling, the score includes a number of musical fragments, gestures, timbres that, obscurely foreshadowed early in the piece, resonate more fully later on. For example, a phrase within the electronic prelude foretells the final, climactic vocal cadence. Or, at a much closer interval, voice and electronics express the horrific realization, "Mir graute ein Gesicht von innen zu sehen" (It disgusted me to see a face from the inside), as if in instant replay. This kind of foreshadowing/backshadowing also crosses languages. When the singer utters in English, "she had completely fallen into herself, forward into her hands," it is just moments after a premonition of the same line has been heard, in the electronic score, *auf Deutsch*. By the end of the piece, the listener will have experienced its main ideas in both directions, in German and in English, on several different but simultaneous musical and psychological levels. The musical logic that motivates *Vis-à-vis* could thus be described as a kind of complex mobius strip.

**Interacting with the Drama: The Real-Time System**

Vis-à-vis is called an "interactive monodrama" because the polyphonic logic of the piece is realized by means of a technology that enables close communication between the singer, the electronic score and the computer-generated video. The computer functions like a virtual performer, as in a chamber ensemble, listening to the changing musical events (for example, the phrasing, dynamics, and register of the singer) and responds accordingly, in its own terms. The point is to capture, in this electronic work, something of the improvisatory spirit that is part of any live performance, and to create a sonic and visual environment in which the singer and computer can interact as a duo, responding to each other's musical decisions.

The electronic score of *Vis-à-vis* is performed using an interactive environment created in the programming language MAX/MSP. This custom MAX/MSP environment produces an audio landscape whose sounds are exclusively derived from the singing voice. A Macintosh computer, running the program, receives audio from a headset microphone worn by the singer. A variety of real-time processes are used on the voice, shifting from scene to scene, according to the tone of the text. Among the real-time processes used are spectral analysis/resynthesis (analyzing and recreating the overtone series of the voice), granular sampling (fragmenting and recombining the voice), harmonization (creating vocal polyphony), frequency shifting (shifting the overtone series of the voice), and envelope tracking (responding to the loudness of the voice). Surprisingly, one of the most difficult parts to bring off was the second long section of the piece —the discursive section beginning with the words: "For example . . ." The real-time electronics for this part track the difference between the singer's consonants and vowels, recombining only the consonants into a percussive polyphony that manages to ampfily the text without obliterating it.

The overall structure of the system invites anthropomorphic allusions, for there are a number of "listener" agents that function throughout: such agents are employed to analyze the pitch, amplitude, and timbre of the voice, while also classifying particular events (such as the incidence of consonants and vowels described above). These listeners

produce data that is used, in turn, to drive other "players"—essentially DSP algorithms—that process and/or synthesize sound. And it is this data that eventually gets passed on to the video control algorithms. Figure 1 maps out the components of the real-time system in the form of a virtual flow chart. (Fig. 1)

FIG. 1: BLOCK DIAGRAM OF THE SYSTEM
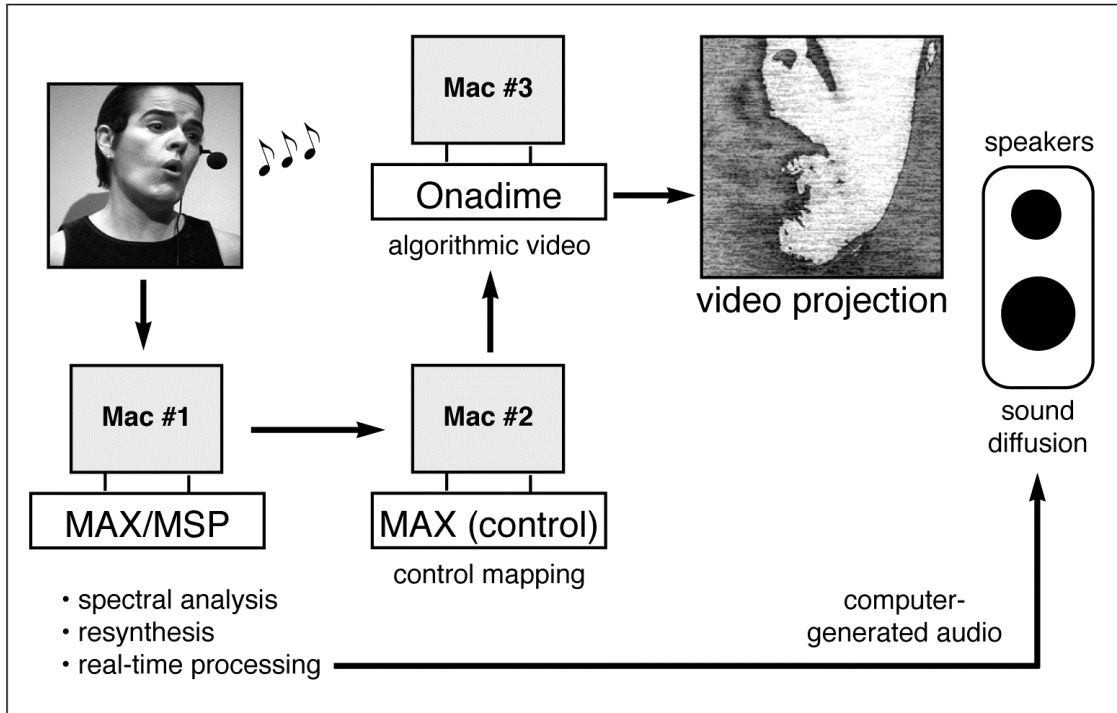
Building a system on this type of architectural paradigm is, in itself, nothing new. However, in the case of *Vis-à-vis,* there is a constant level of "fuzziness" in the algorithmic relation between listener and player. The connections are designed so that the listener agents have a constantly changing degree of autonomy. At times they will choose to listen, other times not. When they are not listening, they have the option of generating their own data, or to do nothing at all. This changing relation between listener and player helps to create a sense of dialogue between human and machine, and gives rise to varying levels of uncertainty—or what I like to call "obstinance"—in the interactive electronics: an obstinance that defines the essence of what we mean by "interactivity." In *Vis-à-vis* such obstinance is manifested at several levels of the system, from those listener agents

that derive data from the audio input, to the DSP algorithms that process and generate the audio, to the video algorithms that control the visuals. One might also say that the live singer adds her own level of obstinance, in the way that she chooses to respond, or not to respond, to the real-time environment.

In concert performance, *Vis-à-vis* actually requires three Macintosh computers: the first runs the real-time MAX/MSP program; a second Macintosh runs a small program (without audio processing) that gives directions (or mappings) to the video computer; and the third runs the video program. The video for *Vis-à-vis* is, like the audio score, controlled in real-time, in response to the events of the different audio "scenes." This visual score was in fact realized using a second custom program—an algorithmic video controller—written in a video programming environment called Onadime. As in the audio score, all the material for the video is derived from raw footage captured mostly of the singer, sometimes speaking, sometimes thinking. The only exception is a single visual "quotation," a lonely woman looking through a windowpane, taken from an image by the turn-of the-century Parisian photographer Eugène Atget. During performance, the video program controls the choice of visual material, their varied combinations (cutting and cross-fading), the rate of change, and the different processing algorithms. But there is, in effect, no video "track" for the piece: in each performance the visual score is created anew.

The Singer's Voice: Reading Rilke as a Performer
As should now be clear, the aim of the multimedia "reading" in *Vis-à-vis* was to suggest the shifting psychological relationship between the subject who narrates the story and the nameless woman who is ultimately the subject of this narration. But this relationship is only fully realized in live performance. Indeed, because the score dictates that the narrator of the piece is also a woman, it is the female voice that forges the delicate relationship between what we might call the two subject positions of the drama, that of the observer and the observed. Our paper concludes, then, with a few remarks from the singer herself, reflecting on what it is like to experience Rilke as a performer.

To sing a dramatic work like *Vis-à-vis* is to face the same challenges one encounters performing any music with text. In performing the vocal part, I am expected to embody the words in some way, to understand not only the character of the story but also her situation. In *Vis-à-vis*, the issue of character becomes interesting, not only because of the frequent changes in "tone of voice"—speaking or singing, English or German—but also because of the vocal range. The part spans over two octaves, and those registers seem to suggest something about the psychological state of the first person who narrates the story, from deep resignation to hysteria and back again.

The electronic score supports the live singing in more than one respect. First, many of the processes and sounds in the piece, as mentioned above, were originally inspired by recording sessions in which I improvised readings of the text, both as recitation and as song. (Indeed, some of the score's vocal melodies were ultimately transcriptions of those earlier, improvised sessions, thus suggesting that, even at a very early stage, the composition of the piece was essentially collaborative.) The total soundscape reinforces the live singing, then, simply at the level of timbre, for its sounds, however modified, are essentially based on my own voice. But the sense of support far exceeds this color similarity. More gratifying for me as a performer is the way the interactive system can actually promote a certain kind of singing, or, at least, modify my behavior. In the second half of the piece, for example, after the crucial turning point ("Aber die Frau"), the system is designed at points to recognize sudden changes in amplitude. If I forget about *bel canto* lines, and exaggerate the discontinuity of the phrase, a marvelous thing happens: suddenly, I am surrounded by a whole roomful of voices, my voices, in fact, keening (like a Greek chorus, perhaps?) in sympathetic horror at the vision I am about to experience. On the other hand, the system's "obstinance" can leave me entirely exposed. In one performance, my phrase might be completed by a clangorous bell-like punctuation, the result of the real-time "listener" hearing, and reconfiguring, the formants of my vowels. In another performance, the system will have turned a deaf ear, so to speak, and my phrase might be uttered, by contrast, in complete silence. The uncertainty has a salutary effect on the performance event. Not only does the changing relation force

me, as a singer, into the present tense, but the organic change in the electronic landscape encourages me to listen to the sounds anew, adjusting my phrasing accordingly.

Of course, it must be emphasized that this sort of organic adjustment is the result of two performers, not just one. For, although it looks like a one-woman show, *Vis-à-vis* is in fact a delicate piece of chamber music. At every turn, I am joined by another player: the composer/conductor, who is seated some distance away, behind a MacIntosh computer. He too responds to the changing musical environment, cueing the next event in the electronic score with a carefully tuned ear, responding to the subtly shifting phrase rhythms. It is ultimately this unseen collaboration between a whole ensemble of listeners and performers—both virtual and real—that produces the drama of Rilke's story, and brings out the many faces of *Vis-à-vis*.