

I Can Just See It Part A

Once many years ago I knew a woman—I'll call her Sandra—who was suddenly drawn to the idea of singing. She had become involved in community theater, was bitten by the performing bug, began taking lessons, and looked for opportunities to perform. Somewhat older than young, Sandra was new to voice performance and I was curious about her motivations. After all, most musicians get started early in life. So I asked her, "Why do you want to sing?"

"Because it's my vision—I can just see myself in those spotlights" was her answer.

Over time, as I came to know her better, I began reflecting on that vision. She didn't say, "I love to sing." She didn't say, "I love music." She didn't even say, "I especially enjoy listening to vocal music, and I want to participate in the pleasure of creating such music—constructing it physically, rendering it artistically, and bringing it to life spiritually." No. Instead, she simply said, "I want to be a singer."

It interested me, this distinction between *being* something versus *doing* something. So I watched Sandra as she developed her performance instincts in "open-mic" gatherings. At first, of course, there was the unavoidable self-consciousness—that weird awkwardness that only sets in when you're actually on stage. The act of performing in front of others always feels painfully different than the way one imagines it in the shower or the car.

Monitor speakers are tricky to get used to. The microphone pops with wind noise, and fluctuates in volume. The audience is nothing but a yellow-blue smear of waxy light. Drum fills emerge off-beat—or at least you could swear they do, because, after all, you were keeping up; and your foot taps are out of synch with his hi-hat, so it must have been him, right? Just as you start to move to the groove, a guitar neck jumps over, hits your shoulder, and then pulls away. "What was that? Is that me out of tune? Oh, now I've forgotten the words."

Performance is *very distracting*.

But as one becomes accustomed to the mechanics of performing, more and more attention can be paid to the music. That's what I was listening for as I observed Sandra. The *music*. Because that's what *singing* is all about.

Unfortunately, Sandra never mastered singing. And the reason became quite apparent over time. She *didn't know how to listen*. She closed her eyes as she bobbed at the microphone, and was obviously unaware of events around her. There was no connection with the musicians or the audience. She was like Tommy—lost in an inner world where she was seeing herself so clearly that there was no attention left over for other performers—unable to attend to music.

She was *being a singer*. But almost certainly would never *learn to sing*.

I often speculate on what is happening internally to produce human behaviors. In Sandra's case, I believe that she constructed a mental image that was primarily visual—which is to say *spatial*—with a strong set of associated feelings that were social, self-preoccupied, and somewhat juvenile (Mommy, look at me!). She did not have a keen sense of music, because music proceeds across time, and Sandra had collapsed time into a single state—a sense of aliveness and wholeness, a feeling of self-ratification. She was fascinated by the *end result*, the product of music, but could not atomize it into the ordered sequence of temporal gestures that make music possible. Sandra was trapped in a spatial vision that disallowed time. She could “just see herself” in the spotlights.

But she could not hear.

I Can Just See It Part B

Expanding on Sandra's plight, there are others driven by vision rather than goals. I know people who “want to be a writer someday.” I had a friend once who said, “When I retire I want to be a painter.” Many others want to be golfers, dancers, socialites, woodworkers, potters, rich, and wise. But they are so busy dreaming about *being* something that there's no time left over to *do* anything—and all of those goals can only be achieved by *competent and intrepid doing*. Without doing, there is no being.

In particular, I am reminded of a young composer wannabe named Dale, who had a vision similar to Sandra's. He could “just hear the music,” clearly, as though it were being played in his head. But when I asked him to sing it, or play it on his guitar, or write it down—he was stumped. He preferred to fantasize about the image in his mind rather than to realize anything that might establish it unambiguously as music and allow him to share it with others.

Like Sandra, I believe that Dale constructed a spatial, visual image of the *result or product of music*, rather than imagining music itself. To test this hypothesis, I asked him to describe what he heard. What he described were not sounds, but feelings.

“I hear distant echoes—almost like the faint ripples of my own birth—that come closer and more demanding. I feel swept up by currents that sometimes push me down, washing over me like waves of sadness, and sometimes lift me up, propelling me toward a glorious sense of wholeness that surrounds me with light and makes me feel close to God.” And so on.

Sounds like a great piece of music, doesn't it? I know *I'd* sure like to hear it.

But what Dale is describing is an *internal mental and physical state*—much like the states that any of us can construct when listening to music (or enjoying similar physical and mental pleasures). When he listens to music, he experiences visual reveries, and he associates them with the music itself. So when Dale mentally reconstructs the reveries, he thinks that he is hearing—indeed composing—music.

Dale basically cuts to the chase. That is, rather than wading through a time-consuming sequence of individual auditory gestures, Dale collapses time into a spatial image that is rife with feeling, and then indulges himself in the reveries associated with that affective state. In other words, it all *appears at once as a feeling*.

I Can Just See It Part C

Referring now to ASR—when people “imagine” what a conversation with technology would be like, they do what Sandra and Dale did. They collapse the image into a visual spatial construction and then imbue it with feeling. This is why speech interfaces seem—on the surface at least—as though they should be natural and delightful. When we develop a “vision” of speech we remove time from the introspection, and instead respond to the single state that is the end result or product of moving through the speech dialogue.

Like Dale, in other words, we *cut to the chase*. Rather than endure step by step the many moments of tension, confusion, wrong turns, and repetitions that characterize real human speech in real live dialogues, we jump to the end state—the affective state in which the conversation is now successfully completed and we are enjoying

the afterglow of the sequence. This is the vision that preoccupies true believers in the speech community.

We should be careful when a design idea for IVR—or for that matter any proposed speech dialogue—appears all at once as a feeling. Because what we are imagining is the end result or product of the dialogue, not the sequence of speech acts that lead to it. A speech dialogue, like music or theater, unfolds over time, and it is *time* that complicates the static afterglow of user delight sought so vigorously by speech designers.

In fact, end users exist within a different psychological set and a different psychological setting than does the designer, the call center manager, the speech visionary, or the focus group participant. What's more, there are many paths through the dialogue—paths not experienced nor even imagined by the designer. And each unique set and setting—what we can think of as the *starting state* of a given user as she enters the temporal sequence of the conversation—leads to and is influenced by the unique path through the dialogue taken by the individual caller.

What emerges at the other end is not one, but a huge array of possible internal states that can reside in the user's mind and body at the end of the sequence. And not all of those states create the afterglow that we envision when we picture success. Indeed, most of them are troubling and discordant. All users who failed at their task, for example, and many successful users who experienced extended periods of tension or uncertainty may be left with an end state that is memorable but negative.

And—if branding is successful—callers associate that state with the enterprise. Careful what you wish for!