

Reexamining "Reexamining Static"

BY J ZEVIN & JIM ELLIS

[see <https://vimeo.com/73070579>, "Reexamining Static: ten minute excerpt of 56 movies playing simultaneously" (2013) by Jim Ellis]

J Zevin: When I first encountered "Reexamining Static: ten minutes of 56 movies playing simultaneously," it was through your Facebook feed. There was some argument there about whether "Born to be Wild" had been somehow pushed up in the sound mix, which you insisted it hadn't. I was drawn to this because it's an unusually clear example of how auditory attention can get captured. In vision science there are actually pretty good models of "saliency" that predict, based on low-level stimulus attributes like contrast and color, what parts of an image will draw your attention. For example, [here's a video](#) of the Itti and Koch model finding road signs in natural scenes.

It turns out that road signs are well-designed to take advantage of the kinds of features that draw our visual attention. We know a lot less about how this works for audition. For vision, some of the problem is decomposed at the retina, where you first make contact with the visual world – two-dimensional space is represented relatively literally as location on the retina, color is encoded by different cells than the ones that do luminance – these channels turn out to be very important to determining where you will direct your attention in visual scenes. Further, we have a great measure of visual attention at work: when something in a scene really grabs your attention, you move your eyes to focus on it.

The auditory system is harder to study. We take sound in as changes in pressure over time through our cochleae, so that things that seem fundamental, like pitch and timbre actually have to get unpacked somehow from the same input. Even the location a sound is coming from is computed by comparing information between the two ears, and taking into account how the shape of your head distorts sound waves. And of course we don't move our ears around to hone in on the signals we find most interesting, so it's hard to measure what part of an auditory signal people are attending to without interrupting them to ask, or giving them a test that involves memory afterward. So, although there's [a long history of work](#) on things like [the "cocktail party" effect](#), the state of the field is still in a place where intuitions based on the phenomenology of unusual stimuli can contribute something. Maybe a lot.

EAR | WAVE | EVENT

Issue Two
Spring 2015
earwaveevent.org

Why, for example, does the average of 56 soundtracks turn – for the duration of the Steppenwolf song – into basically some music plus some background noise? Presumably the other soundtracks contain things that are interesting to listen to, or would attract our attention in other contexts?

Jim Ellis: Well, first I should maybe address that a certain segment of viewers are going to be a bit confrontational when they're presented with 56 layered Hollywood feature films that are played simultaneously. Not only is it challenging to experience, but some folks see it as outright theft, and shame on me for not creating everything from scratch when I baked my cake that night with samples that are an hour and half or longer each. Experiencing 84 hours of feature films in an hour and a half.

I didn't want to influence things too much in this experiment. I wanted the purity of just experiencing a bunch of films playing at once. I made certain overall adjustments to the entire visual mix, but that was more to make the experience less overwhelming. I wasn't just interested in sensory overload, I wanted to experience as much as I could without pain, so I rounded some edges visually (slight expanding delays and feedback softening) to make it more palatable, tolerable, understandable. The audio is completely unprocessed beyond the pile up. I would have even played more films if I had had the computer power to do so in real-time, but that's where I hit the wall. It all stemmed from a real-time visual performance friend of mine wanting to know how many movies I could play in a simple compositing network in real-time with the tools I use and build my art and mad inventions with.

So, back to the *Easy Rider* soundtrack, which was NOT increased in volume. Yet it is the only thing that you can sonically discern with consistency in that section of the film. Why? You hear bits of other things that come into focus like radio static. Why do the other sounds vanish?

Let's start with just that song alone. "Born to be Wild" is a loud, testosterone driven, distorted dirty little ditty that's designed to grab your attention like a screaming alpha motorcycle. It's a repeating pattern (of slight variation) that reinforces itself through instrumental convergence, punctuation, and general human moaning and screaming of stretched out words that rhyme when the song crescendos. It's alive with emotion. And that's just the composition. Then it's completely sonically compressed in the studio to follow the formula for pop songs designed to have more contrasty punch on the radio. The distorted song was recorded at maximum with a signal that was already flattened by pushing against the ceiling... and then the dynamic range compression boosts the quiet sounds. You just end up with

something that is loud. Who knows how the file was dynamically compressed yet again when it was digitized. Could be they just scaled the wave some more and kept it from clipping? Okay, so that's the song in its pure form; the way you hear it in the soundtrack to *Easy Rider* without 55 other movies playing simultaneously. So if we add 55 other movies, we still have the loudest guy in the room. Now here's the question: Is the song alone just loud enough to penetrate everything else, or, what part does our ability to focus on the familiar play?

I'm unsure. Either way, this elicits a lot of questions. How much does familiarity play into listeners' experience? Do people hear "Born To Be Wild" as much if they aren't familiar with the song already? Is this the same mechanism that allows us to hear the first second of a familiar song and immediately recognize it?

In one second's time, we have access to this sonic map that we have partially memorized. If the song is 3:38, we roughly know what's going to happen sonically for the next three minutes and thirty-eight seconds of our lives. If we start to get lost, the repetition, the tonal progression, and the rhymes will help us keep track of what's coming up next. Foot tapping to sync future prediction with body mechanics. All this within a second of hearing the start of the song.

If heard on the radio and punctuated by static, our minds would fill in the blanks or smooth them over like blind spots in our eyes. It's familiar territory. When 55 other soundtracks are playing, maybe we are drawn to the layout of that familiar old road, no matter how obscured. Its simplicity and minimalism are a form of dominance, its loudness compression contrasting reality into solid color blocks with cartoon outlines. Road signs in the salience test, right?

We can't and don't keep track of every bug outside, and at a certain point we smooth out what we need to, and we pay attention to what we think is important at the time. However, if a bee comes too close, we'll pay a lot of attention to that because experience and convention have taught us to pay attention to it.

Now what if you layer things that are designed to excite with things that are designed to elicit a feeling of suspense. Throw in arousal. Maybe some panic. A puppy. Now another excitement. All of these films use similar compositional grammar and story telling (most abide by a three act structure over the period of a standard hour and half), yet they are all about differing subjects. So though there tends to be some commonality within the

ingredients, in the end, it's fairly haphazard. It becomes just interacting shapes. But we still find faces in the clouds. But was that a real face? Because real faces are there, or was that my cloud imagination? All of these movies' intended meanings bounce around off each other and are reassembled in washes of conflicting and re-integrating form and meaning. Form and meaning that become very subjective for the viewer. It's a fight for attention, but it's also dramatic structure and a bit of sensory overload. "Born to be Wild" is an anchor when it plays. When "Born to Be Wild" is over, it's even disconcerting for a moment. Then there's nothing else for long stretches of time except the free association of a Hollywood Rorschach test.

JZ: Yeah, there's definitely something primal about repetition with variation, some sweet spot of predictability that people talk about as making music satisfying.

[Elizabeth Margulis has a nice formulation for this:](#) "repetition invites us into music as active participants."

We are not usually "active" in the literal sense of singing along, or even dancing or tapping our feet. We are "active" in the sense that we are predicting what's coming next. Neural metaphors are useful in thinking about how this might work.

[One idea](#) is that part of the pleasure of music is a balance between having our expectations met and having them challenged. It could be that "Born to be Wild" pops out because it is so well calibrated to elicit predictions about what will happen next from moment to moment.

Familiarity is playing a role, too, though. Consider how hard it is to find the beginnings and ends of words in an unfamiliar language. Our experience of hearing our native language as a series of discrete words is the result of a learning process that is still [an active topic of research](#).

I suspect that familiarity and repetition are interrelated here. To be even moderately familiar with "Born to be Wild" is to be familiar with two or three motifs that repeat with some variation at a time scale of about three or four seconds. That's different from being familiar with the opening sequence of, say, *Withnail and I*, which I'm sure some people could recite by heart with nearly perfect timing, and yet it would likely fail to bubble up to the surface even for those maniacs, because it doesn't reinforce itself the same way as "Born to be Wild" does. On the other hand, there's relatively loud music later

in the piece that probably "pops out" from the rest of the soundtrack when played alone because of its structure, but gets washed out here because it is less familiar, or less repetitive.

JE: We're creatures that abide by a fairly regimented existence, and repetition does seem to be rewarded. We seem to need prediction and met expectation to hold onto, like a tree or house on a spinning ball of dirt... with the occasional mutant gene to keep us from getting too bored. When one first sits down to watch a movie, we can kind of figure out what most movies are supposed to be about within a few minutes, sometimes seconds. In music, you mostly only have seconds before someone will determine if it's something worth devoting their attention to or not. Is their head in the right spot for this? Will it be sexually arousing? Will it piss off my God? Is this what I self-medicate with? Does it help me to better understand how to live my life in a manner that I'm comfortable with? All these genres and movements have their accepted variance within the greater accumulative mode of expression.

Which makes me think of the [neuromarketing experiment with the film *The Good, the Bad and the Ugly*](#) that you turned me on to, J.

JZ: In the original experiment with *The Good, the Bad and the Ugly* – [there is a good non-technical summary, and some thoughtful commentary here](#) – the moments in the film that generated the highest levels of similarity in viewer's brains included moments of dramatic climax, close-ups (generating high degrees of similarity in regions that we know activate strongly in relation to faces), and establishing shots (prompting correlations in areas that seem to encode visual information about locations, etc). [A follow-up study](#) in which data was collected about someone's brain activity while telling a story, as well as from the brains of people listening to the story, found that correlation between the story-teller and her audience predicted how well people understood and remembered the story.

In a way, this is not surprising. Brain-brain correlations just tell us that the movie we're watching is successfully getting people into the same mental state. When that is the primary goal, as in melodrama or a television commercial, we can measure success by how well that works. In communication between people, we can think of language as a way of trying to reproduce our own mental states in someone else's mind. So, the more successfully we do this – either because we are interesting or the person we're talking to is receptive to what we're saying and shares common ground with us – the more similar our brain activity is going to be with the

people listening to us. [We have some data](#) suggesting that this is related to predictability.

It's easy to get mystical about how a good storyteller can cause the brains of her audience to be coordinated with one another, but the scientific challenge is figuring out how that works. On the other hand, the idea of "brain synchronization" a powerful metaphor for communication. Suzanne Dikker, who led the work looking at the role of prediction in similarity of brain responses, also [works on art projects](#) where real-time measurements of EEG are used to produce visualizations and mechanical movements.

JE: As for *The Good, the Bad and the Ugly* experiment, what I'd be more interested in seeing the data for is what body scents were individually and collectively released at what points in the film within the context of a group/shared viewing experience. That to me is more of an interpersonal soup of meaning through interaction. Not because I want to sell them shit or control their minds like some others do, but rather to simply create an artistic progression. That and because I find all modes of communication interesting. Especially one that so often works so subconsciously as scent does. It's almost taboo because it's that animalistic sense that we often deny in the context of human interaction outside of sex.

I really want to make scent songs, to use basic music creation techniques to elicit responses through olfactory performance. There are now methods and tech for scent synthesis. I got to sniff some of that back in 2003 at a convention, but I've never had the chance to play with the tech myself. The machine had the cutest little dispensable sanitary olfactory delivery device covers. The equipment is too expensive though. Back then they had a device that contained a cartridge with 128 "primary odors," which could be mixed to replicate natural and man-made smells. I like the idea of that, primary odors, but I wondered how carcinogenic they were. The company is out of business now. But what are the 128 primary song or film constructs that you weave the moment and individuality around? Are the primary scents of "Born to Be Wild" as basic and in your face as the song is? Maybe something like when that guitar riff happens as blasts of armpit sweating beer pulsate with occasional exhaust and an overall progression of smells from country to city. What are Foley scents versus musical smelltrack? What would an abstract smelltrack be? I think we'll be publicly encountering some of that tech fairly soon. Of course pornographers and fart obsessed 12 year olds are going to get obnoxious with it. Most of the current research is coming out of Japan.

There was Smell-O-Vision back in the 50's where scent was released into theater ventilation systems at key moments in the film. [Historically, it is understood to have been ridden with distribution difficulties](#). Scent delay to outer reaches of the ventilation system. Either too little, or too much. That and lingering scent just piling up on itself because it's too difficult to get the right airflow to cleanse and control the pallet.

JZ: Whoa.

Smell is even more nebulous and difficult to get a handle on than sound. Aside from the practical problems of clearing odorants from a room, I think the temporal dynamics of the experience of smell are intrinsically hard to control because of the way odorant molecules find their way to receptors, and the way those responses evolve over time. It's not like a flash of light hitting your retina, or a sound wave impinging on the cochlea. If you were going to make music out of smells, it would have to be slow, maybe so slow that you'd have to broaden your definition of "rhythm" to even think of it as music. And you'd have to be careful about repetition, because [the olfactory system adapts fairly dramatically](#). Listeners might have to be careful about listening to the piece too many times, or risk habituating to the odors and lose the ability to smell them at all.

Actually, so many things break down when you try to think about making music with smells. How would smells relate to one another, so that you could get something like intervals from sequences or chords? What would it mean to have "perfect pitch" for smells? (Although, like perfect pitch, [which is more common in populations with tonal languages](#), there are languages with tons of smell words, and [speakers of these languages are better at identifying and discriminating smells](#) than the rest of us)

I guess all of this gets challenged in "Reexamining Static," too...

JE: In some ways "Reexamining Static" reminds me of lingering smell build up. The nose is quick to become exhausted by strong odor, it just kind of shuts down its sensitivity due to exhaustion. I haven't yet made it all the way through the hour and half of "Reexamining Static," after a point you just have to let it wash over you.

JZ: I haven't been able to finish the ten minute clip in a single sitting, although I think I've seen all the parts.

JE: So where does roughly 56 channels of information being experienced simultaneously actually evoke comfort? Nature, which makes sense because we are born from it and have evolved to function within it's continuous streams of input. Where else? Music, right? Orchestras, for sure. They have to know in advance what purposely related fragment of the larger global pattern that they are going to perform. They have to practice their part, and their interaction as a group machine. We also get it in 200-channel overproduced Pro Tools generated popular music. You might have twenty channels alone just for the word, "baby." But the density of each such channel is often variable and also quite sparse, whereas with "Reexamining Static," it's constant in each channel. It blurs together out of the sheer quantity of information. Like having lunch in a crowded convention center cafeteria, with a thousand people, and yet one still manages to spot their missing friend across the room.

Where it differs is that the multiple streams/channels of sensory information are more diverse in subject matter and have shifting conflicting/reinforcing spatial cues. I'm really interested in upping the number of informational streams that can be comfortably perceived without the overall effect becoming too nonspecific and/or too reduced. How can continuous data streams be more harmonic, and where does depth come into play? I see multimedia as just starting to have an effect on the way we learn. Has music served as a place holder for new multi-channel modes of communication? When film first appeared, and then again with 3D in the 50's, people used to flinch when something on the screen was moving toward them.

Now, when I went to see the second *Transformers* movie to study its effects, I had a non-stop barrage of explosions and mecha-morphs, panoramic doppler, and I fell asleep. Without pauses in the action, the entire thing just became a wash and was utterly boring. I just tuned it out. It reminded me of an extended car trip I took with my cat. The first day the cat was unable to handle the input, was having a breakdown. After the cat slept for a while, he had adapted and was fine. Maybe I should watch "Reexamining Static" for a couple days and see what happens.

Jim Ellis is a multidisciplinary artist and pioneer in the field of realtime performance animation. His work is/has-been showcased in numerous museums, galleries, universities, as well as ACM1, and Siggraph. He has worked with the likes of Rush, Terrence Malick, MTV, and has been faculty at CalArts, and Loyola Marymount. Jim is featured in "CGI: the Art of the 3D Computer Generated Image," as well as having contributed articles to the New York Times, the L.A. Times, and various magazines. Jim is currently developing new forms of adaptive computer interfaces for both creative and scientific use.

A partial selection of his short films can be viewed here: <https://vimeo.com/album/1877924>

J Zevin is Associate Professor of Psychology and Linguistics at the University of Southern California. He studies the perception and comprehension of spoken and written words, and how those processes unfold in the brain. Recently he has been grappling with the roles of context and prediction, and has begun nourishing nagging doubts that the way these topics are studied in the lab denatures them in important ways. Visit his lab online at: <http://zevinlab.org>