Machine-aided close listening: Prosthetic synaesthesia and the 3D phonotext

Chris Mustazza <mustazza_at_sas_dot_upenn_dot_edu>, University of Pennsylvania

Abstract

This paper proposes a new listening methodology that lies, in some sense, between Charles Bernstein's close listening and Tanya Clement's distant listening. While it's tempting to think of these approaches as presenting an integral of scope, Bernstein's work proposes the use and tuning of the human ear for understanding the sonic materiality of the phonotext, while Clement's work, following its namesake of Moretti's distant reading, calls for a form of surrogate listening — using the machine in place of the ear. Building on Bernstein's work and toward Clement's, I suggest Machine–Aided Close Listening as a methodology that uses data visualization as a prosthetic extension of the ear. This is to say that by considering three dimensions of the phonotext — 1) the textual manifestation of the poem, 2) the audio of the poet performing it, and 3) a visualization of that audio — the reader–listener can empirically confirm the ear's impressions and expose new facets of the sonic materiality of the performed poem. In support of this methodology, I will demonstrate a new digital tool for aligning these three facets of the phonotext.

Any three-dimensional object, which we see dispassionately, is a projection of something four-dimensional, something we're not familiar with. (Marcel Duchamp, *Dialogues with Marcel Duchamp* [Cabanne 2009, 40])

How might technologies for pre-treating and graphically presenting phonotexts supplement our practices of "reading" them? And in what ways might they come to transform those practices? (Steve Evans, "The phonotextual braid")

Introduction: Perceiving the materials of poetry

In recent years, critical inquiry into the craft of poetry has taken on a particularly material approach, working to catch up with poetry itself, which has long been occupied with its own materiality. From Jerome McGann's *Black Riders: The Visible Language of Modernism* [McGann 1993] through very recent publications, like Bartholomew Brinkman's *Poetic Modernism in the Age of Mass Print* [Brinkman 2016] and Carlos Spoerhase's "Beyond the Book?" [Spoerhase 2017], the visible, typographical aesthetics of poetic form have been shown to be an extension of the poems' content, and vice versa, to borrow from Robert Creeley. Charles Bernstein's *Close Listening*, his intervention calling for attention to be paid to the performed poem, rather than its written manifestation alone, was the logical next step — to recognize the sound of poetry as a dimension of its materiality. For Bernstein, the performance and all of its aurality is not an adjunct

to a primary, written text. He proposes that it is crucial "to overthrow the common presumption that the text of a poem — that is, the written document — is primary and that the recitation or performance of a poem by the poet is secondary and fundamentally inconsequential to the 'poem itself'" [Bernstein 1998, 8]. Part of the necessity in Bernstein's call for attention to the nuanced contours of the poetic performance is an awareness of the materiality of the poem. And when I say materiality, I do mean the material archival object, which includes tapes, records, etc. — but also the sonic materials from which the performance is wrought, including pitch and timing of the voiced language. This sonic form is, of course, inextricably bound to visual form of the poem on the printed page (when a printed form exists). In the time since Bernstein edited *Close Listening*, there has been a resultant energy directed toward understanding poetry performance, but most studies that take the performed poem seriously tend to privilege it over the printed texts of the poem, inverting the problem. It is this play of sight and sound that I propose we examine further. Specifically, I would ask what media archaeological affordances computational technologies provide toward examining the sonic dimensions of the performance of a poem — without discarding the visual poetics of the printed page. What we need is a methodology that accounts for both sight and sound by further complicating the distinction and making use of the wealth of poetic sound recordings available to us.

In this essay, I propose a new reading-listening methodology I term *Machine–Aided Close Listening*. While I will situate this methodology as an extension of Bernstein's work and apposite to Tanya E. Clement's concept of distant listening, for now I will start with a brief definition. *Machine-Aided Close Listening* aims to juxtapose, without necessarily harmonizing, three dimensions of the poetic phonotext: the text of the poem, a performance of it, and a visualization of the audio of that performance. These dimensions form the 3D phonotext. In making sound visible, a reader–listener can discuss aspects of the sonic form of the poem as an extension of its textual content. In support of this methodology, I offer a new digital tool that aligns these three dimensions, available on the PennSound website. [Creely 2017]

2

The tool brings together some aspects of the visual form of the poem (recognizing that the process of digitization of the text is itself a remediation that recreates), including lineation and spacing, with a visualization of the included audio. The visualization, in this case, consists of a pitch curve of the poet's voice, a waveform for looking at loudness and spacing, and tempo stats (the latter of which I will not explore in depth in this essay). These are not the only ways to visualize Robert Creeley's 1966 reading of "I Know a Man", but they offer a starting point.

3

So, in other words, Image, Music, Text — to borrow from Barthes. It is my hope that this tool will create the space necessary for "the grain of the voice" to become a dimension of the grain of the poem, the embodied and physical conditions that make it possible. But I also hope to move beyond the body, toward a posthuman media studies championed by scholars such as Wolfgang Ernst. Ernst writes:

4

> Synesthetically, we might *see* a spectrographic image of previously recorded sound memory – a straight look into the archive. This microsphysical *close reading* of sound, where the materiality of the recording medium itself becomes poetical, dissolves any semantically meaningful archival unit into discrete blocks of signals. Instead of applying musicological hermeneutics, the media archaeologist suppresses the passion to hallucinate "life" when he listens to recorded voices. [Ernst 2013, 60]

I will not go so far as to refuse any kind of poetic absorption and see the performance as pure signal. But what Ernst's work shows us is that when machines write the content, as they do with sound inscriptions, we need other machines to decode the archived material — and not just to play it back, but to *read* it. As Ernst argues, in the end, the machines become the media archaeologists. What I propose here is that we synaesthetically and prosthetically extend our perception of the inscribed phonotext by including the ability to see sound, aided by the machine. Though at the same time, it's crucial to see this research as respecting the poem and its literary historical framing. *Machine-Aided Close Listening is* a tool of hermeneutic analysis — not an exemption to encountering the poem as a human-readable, aesthetic object.

Finally, before getting into the methodology itself, I want to address why an online tool is necessary for this task, and some further affordances of the tool. One of my primary interests in this work is pedagogical. I am interested in what

5

happens when people come together to collectively make meaning of a sounded poem. We have likely all had the experience of reading poetry aloud in a classroom and working through a collective understanding of the poem, while each person looks at his or her copy of the printed work. We may have had the privilege to do this alongside someone playing recordings of the poet reading the same work. But now, for the first time, a group will be able to use a visualization of the performance audio to make collective meaning of the poem. Does the poet read faster during this part of the poem? Does the pitch of her voice tend lower in this particular line, and if so, how does that affect the line's meaning? Coming together to read sound will offer new avenues into our understanding of poetry. This is what makes the tool distinctly different than looking at waveforms or spectrograms generated by desktop programs like Audacity — we can all look at the same visualization at the same time and cite the same material facets. And I believe this research could be applied to work outside of poetry proper. Performances of prose works are under-studied, to be sure. And what about political speeches and the aesthetics of popular motivation? While these examples do not contain visually formatted content in the same way poetry does, I believe we could learn from bringing them into our tripartite formation of *Machine-Aided Close Listening*. But for now, we will work to understand how this methodology applies to poetry and how to situate it within existing approaches to studying poetry audio.

Machine-aided listening vs. surrogate listening

If close listening calls for attention to the minute resonances of sound, operating at a proximity Franco Moretti might call "theological" [Moretti 2013, 48], on the other end of the spectrum (should we want to set these two things at odds) is Tanya E. Clement's notion of distant listening, the meaning of which has evolved over time. Clement said of distant listening, as she was developing the concept:

[6]

> The opportunity to investigate significant patterns within the context of a system that can translate "noise" … into patterns, within a hermeneutic framework that allows for different ways of making meaning with sound, opens spaces for interpretation. This work allows for distant listening. [Clement 2012]

But the term has drifted, as with Clement's more recent work on the HiPSTAS project [Clement 2016], to align closely with the use of high-performance computing technologies to "listen" to thousands of hours of audio in a condensed period of time, in order to make claims at scale, across an archive or corpus. Much like Moretti's distant reading, distant listening has come to suggest a methodology of large scale. For example, Clement and McLaughlin have used distant listening to ask the machine which poets get the most applause, a question that relies both on scale and quantification of the extra-lexical [Clement and McLaughlin 2015].[1] With the proper metadata, we could ask another question Clement has proposed, whether there is a differential in applause for male poets versus female poets [Mustazza 2016]. Or we could ask if we can determine a performative commonality endemic to First–Wave Modernist poets, and how this differs from those in the Second Wave. And so it is clear how a prima facie comparison of Bernstein's and Clement's theories would suggest that they are opposite poles of a spectrum: while close listening privileges the nuanced facets of the performance, distant listening (as it has come to be used) culls through massive amounts of information and suggests synchronic or diachronic patterns. The obvious question, then, would be to ask what methodologies lie within this integral of close to distant. But to do so highlights that the two modes do not operate on a sliding scale: it's not so much the question of scale, or even distance, that separates the two, but rather one of which perceiving organ is deployed in pursuit of meaning-making — and to what degree is/are those organ(s) technologically extended.[2]

One of the most provocative lines in Franco Moretti's call for distant reading is when he claims that, in opposition to the "theology" of close reading, "what we really need is a little pact with the devil: we know how to read texts, now let's learn how not to read them" [Moretti 2013, 48]. If we want to perceive literature at small and large scales, in other words, we should let the computer read for us, and "if between the very small and the very large the text itself disappears, well, it is one of those cases where one can justifiably say, Less is more" [Moretti 2013, 49]. If we're willing to take this claim at face value, it suggests that the machine is not an extension of the eye; it's a surrogate for it. Moving back to listening distance, the same is true. While Bernstein's work calls for the deployment of our ears (in addition to our eyes), Clement's recent work utilizes a delegation to a surrogate ear, one tuned for a very different kind of listening. A question this raises is to what degree, if at all, is a surrogate ear prosthetic, or an extension of the ear. I would suggest that

[7]

distant listening is no more an extension of the ear than is distant reading a prosthetically enhanced eye. In both cases, a surrogate device is deployed and named with a bodily metonymy that allegorizes/humanizes the interface between (phono)text and processor. This differential fractures the spectrum of close–to–distant listening methodologies, setting them as apposite, but not dichotomous. In this essay, I would like to ask the question of how to advance our ability to perform close listening through the use of technological prostheses — the process I have termed *Machine-Aided Close Listening* — while staying in the realm of augmented listening, rather than surrogate listening. One assumption here is that we hear aspects of the sonic materiality of the performed poem and understand them impressionistically, and that these impressions can be confirmed empirically through digital tools, presented alongside the text of a poem. In addition, the machine can reveal dimensions of the poem that are imperceptible (or difficult to perceive) by the human ear alone. These revelations can be used to challenge initial readings and complicate understandings of the poem.

Building upon the several projects that have very successfully approached exploring the poetic phonotext through technology, including Clement's and the work of Marit MacArthur (more on this later), we need a methodology that maintains our attention to the phonotext without discarding the visual semantic aspects of the printed poem. Lineation and spacing are crucial dimensions of the poem and its multi-sensory existence, and these are often unaccounted for in the application of tools meant for linguistic research to the study of poetry. Conversely, there are a number of tools for studying sounds encoded in the text of poems, tools like Poemage[3] and Poem Viewer[4] that focus on augmented close reading. These tools are valuable for gaining an insight into the encoded sound of poems, but when sound recordings are available, I follow Bernstein in suggesting that they must be included in our phonotextual analysis of a poem.

Returning to the idea of the sight-sound interplay so crucial to Modernist poetics, I would note that a central interest of certain strains of poetic modernism, especially for Second-Wave Modernists Charles Olson and Robert Creeley and Afro-Modernist James Weldon Johnson, is the page as a (loose) score for performance. In these instances, line breaks can be performance instructions (e.g., the insertion of a line break to suggest a musical rest in performance), in addition to devices to visually separate, emphasize, or create a visual syntax that is distinct from (though related to) grammatical syntax. To borrow from Kittler's application of Lacan's Imaginary, Symbolic, and the Real, wherein he places text in the symbolic order, bound by a fixed set of characters, and phonography in the Real, the machine making no distinction between signal and noise or what "content" it captures, *Machine-Aided Close Listening* lets us see the collision of the Symbolic and the Real and how Kittler's distinctions are not as clear cut as they seem. And so an alignment of the visual field of the printed page with the performance is absolutely necessary, and what I aim to demonstrate here via the new digital tool. Our aim will be a prosthetic extension of the ear, via the machine and connected to the eye.

Realignments

In 2011, PennSound released the PennSound Aligner, a basic tool to align the text of a poem with the audio of the poet reading it [Creely 2011]. These kinds of alignments had long been the purview of linguists, interested in topics like transcription and phonetics. But the formal/aesthetic dimension of the printed poem requires a faithful (insomuch as remediation can be faithful) treatment of text. Playing "I Know a Man"" through the PennSound Aligner, which highlights the printed words as Robert Creeley performs them at Harvard in 1966,[5] presents what Steve Evans has termed the Phonotextual Braid: a weaving together of timbre, text, and technology [Evans 2012]. The technological thread of the app, here, twines together with text and timbre, sight and sound, and elucidates what would remain obfuscated by exploring these dimensions separately. Specifically, we can see and hear Creeley's use of the page as score. A reading of the text alone might suggest that a reader should read straight through the enjambment, allowing himself to be sonically guided by the punctuation or grammatical syntax. But Creeley stops at all of his Williamsesque line breaks, both sonically and visually highlighting the phrasing of the poem. What we are hearing is Creeley's implementation of Charles Olson's "Projective Verse", which describes how poetry moves from "the HEAD, by way of the EAR, to the SYLLABLE / the HEART, by way of the BREATH, to the LINE", and praises the advent of the typewriter because "[f]or the first time the poet ha[d] the stave and the bar a musician has had" [Olson 2009]. In short, the Aligner is a technology that exposes technology, the mechanics of the poetics at play and their means of mechanical reproduction.
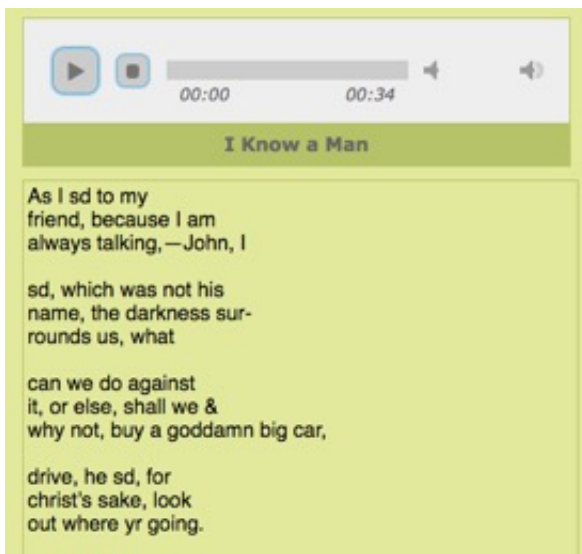
Figure 1. PennSound Text-Audio Aligner

But the current PennSound Aligner is also limited in the ways it can make facets of a poem visible/audible. For example, my unaided hearing of the poem suggested that Creeley emphasizes the "I" of the poem, introduced in the first line by speaking it with a kind of emphasis. I also heard Creeley dragging out the phrase "always talking" to enact the duration of the reflexive exploration of the I's subjectivity, as well as its intentional tediousness for John or other listeners, including the listener to the poem. Too, the line-wrapped "sur-/rounds" also registers with a protracted performance, the meaning enacted in a snake-like coiling around the speaker, the listener, and the line break, the language's sonic and visual materiality circumscribing what can be seen in the darkness and what can be done to stop its centripetal collapse. And what about the characteristic tremulousness of Creeley's voice in the performance? This must surely be read as part of the material of the poem, his trying utter the unsayable of the void that the poem suggests is encroaching. The quavering (and wavering) performance of "why not, buy a goddamn big car" sonically subverts its semantic statement of (ironic) surety of action, its modulating pitch suggesting an ambivalence about whether there is anything to be done besides consume while being consumed. This reading of the poem is supported by the discussion in the episode of PoemTalk dedicated to "I Know a Man", a discussion hosted by Al Filreis and featuring Randall Couch, Jessica Lowenthal, and Bob Perelman [Filreis 2009]. Filreis notes that his reading/hearing of the poem is shaped by the semantics of the quavering of Creeley's voice. He claims that he reads this as symptomatic of the fear of indeterminacy, and lingering anomie. Couch agrees with Filreis, and suggests that the poem must be experienced textually and aurally simultaneously to make this particular meaning. While my close listening of the poem seems to agree with this critical engagement, it is difficult to ground these readings in any material beyond an impressionistic close listening (which I do think is a valuable methodology). We can see the linebreaks and agree where they fall, but the same has not been true of these sonic extensions.

11

It is not my intent to denigrate the sophisticated machine that is human auditory perception. What I would like to propose is that empirical tools can complement — extend — the human ear and lend credence to our first (and subsequent) listening(s). In pursuit of this proposal, I present to you a prototype of a new version of the PennSound Aligner, one that adds a third dimension: a visualization of the audio of the performance [Creely 2017]. In seeing sound, the reader–listener will be able to empirically corroborate perceptions of the material conditions of performance, including pitch, duration, and loudness. To refactor a line from Bernstein's *Close Listening* — "Having heard these poets read, we change our hearing and reading of their works on the page as well" [Bernstein 1998, 6] — perhaps machine-aided close listening will mean that having seen these poets' voices, we will change our hearings of the recordings and our understandings of the works on the printed page, too.

12

The prototype of the *Machine-Aided Close Listening* Aligner, which I developed with Reuben Wetherbee and Zoe Stoller, visualizes the audio of a poem as the summation of a pitch trace (generated by the open source program

13

Praat[6]) and a waveform (generated by the JavaScript library WaveSurfer.js[7]). The alignment with the text of the poem is accomplished using Elan,[8] the open–source annotation software developed by the Max Planck Institute. The alignments, for now, are done by hand in Elan, including mark up for words, lines, and stanzas, which generates an XML file used to render the text of the poem and align it with the audio and visualizations. I'm also glad to say that the custom programming work that Wetherbee did to extend Wavesurfer.js was submitted back to the project, which is licensed under Creative Commons, and was accepted, meaning that others may now benefit from the code developed for this project (free and open access is crucial to what we are doing here).

The intent of the visualizations is to give a relative sense of the pitch dynamics at the line level, and to some extent, across the poem. The specific pitch values can be downloaded for examination and further study, as can the audio file from which they were generated. As with all visualizations, an array of subjective decisions and tradeoffs contribute to the construction of this view. For one, we decided to normalize the width of the visualizations of each line. While this can look misleading, suggesting that each line is temporally equivalent, the alternative would have been a ragged set of widths, which for some poems (think William Carlos Williams) could contain a single word that takes a fraction of a second to read. In order to avoid such a vertiginous experience, we chose a standard spacing and included the start and end times of the segment (from which the duration can be deduced). Other decisions with implications for the way the phonotextual braid is weaved include the placement of the text alongside the visualization (rather than below it), the shading to delimit stanzas, etc. Needless to say, this kind of visualization may not suit every poem in the visually variable canon of poetry, but I do think it offers a place to start.

Figure 2. Machine-aided Close Listening Aligner

And as a way of making this start, I suggest we return to "I Know a Man" (see Figure 2). I had conjectured that the "I" of the poem in the first line was sonically emphasized. Examining the waveform for the first two words of the poem, what would textually appear to be spondaic ("As I") looks more iambic, with a hard emphasis on the "I", engaging with the poem's critique of the speaker's self-centeredness.[9] Proceeding to the poem's famous "sur-/rounds" verse, "verse" here bearing too its etymological meaning of "turn", as in to turn the line break, the reader-listener can easily see the protracted performance of "rounds", as compared to the other monosyllabic words on that line ("us" and "what"). In addition to the unfurling of the sound in its semantic mimesis, the pitch curve for "rounds" displays a rounded-off performance. The pitch starts relatively lower, raises to a plateau of a roughly parabolic structure, and then returns to a lower pitch. Perhaps it's an over-read to suggest that this symmetric wave of pitch enacts the semantic meaning of "rounds", but it can be said to be the representation of a dynamic the listener can clearly feel, using the unaided

technology of the ear. A cursory listen displays something interesting happening with the pitch and duration of that word, and the visualization depicts what that looks like.
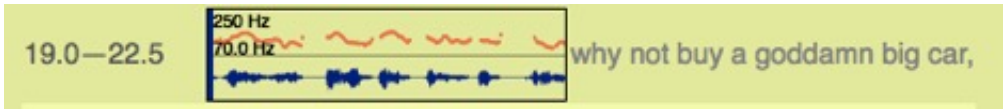
Figure 3. "Why not buy a goddamn big car", visualized

Next, I would draw our attention to the pitch curve for what could be said to be the antecedent to the poem's volta, "why not buy a goddamn big car" (see Figure 3). Unlike some of the other lines, we can see a pretty defined set of frequency modulations in the line, representing the characteristic tremulousness of the performance. If I were one to over-read (and I am), I might say that we could make a new kind of meaning by suggesting that the oscillations in the geometry of the line representing the pitch of the poet's voice enacts the indecisiveness of the speaker I (and the PoemTalk participants) previously recognized. Or, if that seems like a stretch, a more measured stance is that Creeley's voice has a quavering sound to it in this line, and our machine-aided close listening exposes some of the sonic materiality of that sound. And finally, I would invite us to look at the last stanza (see Figure 4), where John is finally afforded the capacity of speech (even if he only exists in the description by the poem's speaker — perhaps I should call him "Creeley's John", as we would say Plato's Socrates). John's pragmatic rejoinder to the speaker's existential dilemma is to demand an attention to the present. Emerging in the last stanza, John's Emersonian pragmatism is rendered as a constellation of sight-sound symmetry. Each of the lines in the tercet is roughly equivalent in textual length, but it turns out that they are also roughly temporally equivalent in their performance (about 1.5s per line). A closer look also confirms the impression of the ear, which is a staccato performance that spaces the words somewhat evenly across the lines. John is much more measured than the poem's "I", whose anxieties drive (pun intended) him to speak lines like the relatively sprawling line "why not buy a goddamn big car". Too, John interprets the score of the page even more literally than the speaker. If the "I" can only perceive in lines, John's ability to focus at the word level, to literally interpret the pauses between words as the speaker does between lines, enacts his attention to the present moment. He is as deeply focused on carefully crafting the sonic materiality of each word as he would be on carefully navigating the car, if he were the one driving. John's (if that is his real name) presence in the poem is then measured symmetry, lexically, aurally, visually. This symmetry enacts the PoemTalkers' reading of John as a force of realism or temporal grounding. Here we see close listening hermeneutics extended and confirmed by our prosthetic synaesthesia and machine-to-machine media archaeology.
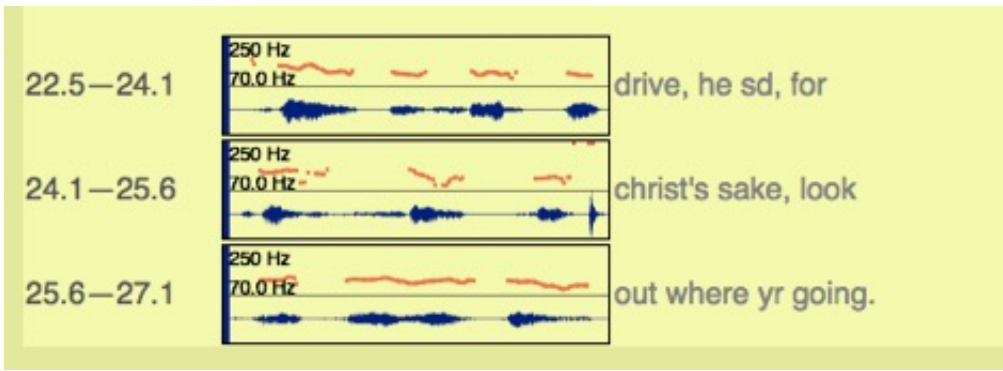

Figure 4. Final stanza of "I Know a Man", visualized

Creeley's "I Know a Man" lends itself well to this kind of visualization, because of the aforementioned Projective Verse poetics that render sound to the page so literally. As machine-aided close listening is applied to other kinds of poems, especially those where the poets read through the enjambment, an extended vocabulary will become necessary. The tool could potentially reveal new organizations in the performed poem, like syntactical units I think of as aural phrasal units, or a grouping of words performed together though not visible in the visual or grammatical syntax of the poem. It

will certainly be a test of this kind of visualization to move away from cases like the Creeley, where the text is in some sense a score for the performance. But it seems best to start with a case that lends itself to this kind of analysis before proceeding to others.

I will just mention here briefly that a future application of the Aligner may be with the sermon-poems of James Weldon Johnson, those published in his 1927 *God's Trombones*. Over 20 years before Olson's publication of "Projective Verse",Johnson was using the printed page as a sonic score. In his effort to preserve the dialects used by African American preachers of the early 20th century, Johnson recorded the sounds to paper. He wrote of his poetics:

> The tempos of the preacher I have endeavored to indicate by the line arrangement of the poems,
> and a certain sort of pause that is marked by a quick intaking and audible expulsion of the breath I
> have indicated by dashes". [Johnson 1927, 10–11]

. In other words, Johnson is using the visual space of the paper to encode sound: line breaks carry a rest duration and em-dashes represent the sonic contours of breath as rhetorical punctuation. At the time of the composition of *God's Trombones*, Johnson did not know that he would have the opportunity to make sound recordings of the performances, which he did in 1935 [Johnson 1935]. Given the inextricable sight-sound bond in these poems, exploring them with the Aligner offers the opportunity to explore sound, scored to the page, performed back into sound, and then rendered legible through these visualizations — sight-sound en abyme.

Following on Marit MacArthur's groundbreaking *PMLA* essay "Monotony, the Churches of Poetry Reading, and Sound Studies", there is much room for a kind of machine-aided close listening that's less close (farther?) than what I have been performing. MacArthur locates the so-called poet-voice of the academic poetry reading, a sonic dynamic she refers to as "monotonous incantation" in a particular pitch dynamic, that of reading each line starting at a higher pitch, and descending toward a lower pitch that correlates with the end of a line/aural phrasal unit. She argues:

> What many contemporary academic poets have inherited of that style is a subdued strain of
> monotony without progressive intensity or the counterpoint of audience participation, a distinction
> that mirrors some of the differences between a mainline Protestant service and an evangelical one.
> [MacArthur 2016, 59]

As a counterpoint to monotonous incantation, she explores the hieratic dimensions of recordings by Martin Luther King, Jr. and suggests that while these too are inflected by sermonic overtones, they evoke different religious cultures. She writes:

> Gradually rising in intensity, King's voice also begins to use a wider pitch range, reaching higher
> and lower as it becomes more emphatic. What the graph does not reveal are the cuts, or
> interjections, from the enormous audience (audible in the recording). [MacArthur 2016, 57]

MacArthur uses a reading-listening method exactly like what I term machine-aided close listening to read both the sounds *and* the silences of the speaker, the figure and the ground.[10] Let's return to the James Weldon Johnson recordings. Johnson's poems and their strict scoring of sound to the page, like the Creeley poem, offer the opportunity to explore whether these same sermonic dimensions exist in Johnson's mimesis of the sermon. But they also reveal that we should be reading the silences in the recording. It occurred to me, after reading MacArthur's essay, that Johnson's interpretation of the line breaks is very different than the Second-Wave Modernist poetics Olson would come to propound. Johnson was not trying to model a compositional approach where the poet's "hearing" of a phrase is transduced (and translated) to the page. The spaces at the end of each line are left empty for audience participation. If these poems were presented as sermons, the congregation would have the opportunity to react to each line in what MacArthur refers to as "cuts", spaces for audience participation and the response to a call. Because the recordings are made in the sterility of a studio environment, there is no audience to respond. But the silences are left as audible signifiers of absence. This sonic fort-da reaches toward Craig Dworkin's call "to linger at those thresholds and to actually read, with patience, what appears at first glance to be illegibly blank" [Dworkin 2013, 33]. Dworkin's use of "threshold" is very much an invocation of Gerard Genette and his exploration of "seuils", thresholds or paratexts

[Genette 1997]. In this case, our machine-aided close listening explores that the trailing silences of each line and presents them as thresholds to an imagined audience, leaving the reader-listener to fill in the proper response to each call.

As these readings of the Creeley and Johnson poems demonstrate, machine-aided close listening allows for analysis at different scales within the poem. But it also may build a bridge to distant listening methods. In order to perform the kind of work done by Clement, to use a computer to scan through many hours of audio in search of patterns, the poetry audio scholar must have an in-depth sense of what sound looks like/sounds like to the machine. This comes from hours of staring at spectrograms and pitch traces to develop a descriptive lexicon of sonic dynamics with which to teach the machine. To search sound with sound, to say "show me something that sounds like *this*", we must be able to articulate, in minute terms, what "like this" means. Unlike a Google search, which to some degree returns text that is identical to the search parameter, distant listening is a matter of tuning and vantage point. It is a process of translation, with the full weight of any poetic lingual translation. Machine-aided close listening provides the ability to learn the grammar of sound, which is at the threshold to any phonotextual analysis, and weigh it at scale. [20]

Conclusion: Poem Lab

I don't mean to suggest that the visualization I've chosen to start with is the only way to explore the sight/site of sound. Far from it, this is just an opening volley for a set of visualizations in a new section of PennSound called "Poem Lab". Future visualizations could include spectrograms, and even overlays of visualizations of multiple performances of a poem, to return to Bernstein's claim about the multiplicity of the poem. For example, for William Carlos Williams' "To Elsie", we could make a comparison of all eight readings of the poem that exist in the PennSound archive [PennSound n.d.]. Does Williams always read with the same phrasing and pitch dynamics? Are these dynamics hallmarks of the literary scenes he was a part of (i.e. inflections learned from attending other performances), or are they particular to this poem? How might machine-aided close listening help our understandings of poems that tend toward sound poetry, yet still have a textual dynamic? I'm thinking here of Amiri Baraka's "Black Art", and lines like "Airplane poems, rrrrrrrrrrrrrrrr/ rrrrrrrrrrrrrrr . . .tuhtuhtuhtuhtuhtuhtuhtuhtuhtuh/. . .rrrrrrrrrrrrrrrr . . .", sound symbolism that emerges from an otherwise semantically regular body of text [Baraka 1966]. And all of this is to say nothing of the rich analyses that could come of aligning videos of certain poetic performances with their respective texts, sounds, and sonic visualization. At the "Les archives sonores de la poésie" conference that took place at the Sorbonne in 2016, I saw Gaëlle Théval give a talk on Bernard Heidsieck's performance poem "Vaduz". Crucial to the apprehension of the poem is seeing Heidsieck manically flip through a legal pad from which he reads the polyphonic poem. One could imagine an application, like the aligner I present here, that allows for highlighting portions of the pitch curve/audio visualization and seeing the corresponding video for the section while it plays. To some extent what I'm proposing here is the recuperation of a poetic mise-en-scene, in the way Artaud would use that term, which is to say the decentering of the privileged semantics of speech, and a reclamation of a rhetoric of the stage. Artaud famously denounced the use of dialogue in plays, arguing for a visceral, sensual poetics articulated through a mise-en-scene predicated on the presence of the body in performance and reception. Or, as Charles Bernstein notes, "the poetry reading is always at the edge of semantic excess, even if any given reader stays on this side of the border" [Bernstein 1998, 13]. [21]

Machine-aided close listening is the next step toward attempting to apprehend more of this semantic excess. I don't think it would be excessive to say that the poetry audio archive is fractal in nature, in that each document in the archive is itself an archive, a collection of rich semantic information fraught with fissures, aporias, and absence. We venture into the phonotext on an archival quest for origins, to get closer to the moment of composition and meaning by approaching the already mediated bodily instantiation of the poem. The ever-present epidemiology of Derrida's archive fever notwithstanding, machine-aided close listening can provide us with the ability to read the material facets that are in some sense part of, and in another way symptoms of, the semantic excess Bernstein identifies in the performance of poetry. As an early modernist might use a Hinman collator to compare versions of a text, as a prosthetic extensions of the cognitive ability to identify difference, so too should the poetry scholar work to extend his or her senses to navigate the sensory archive of the phonotext. [22]

Notes

[1] Note Bernstein's comparison of "extra-lexical" to "extra-semantic" [Bernstein 1998, 5].

[2] See Marshall McLuhan's "The Medium is the Message", where he argues that media are extensions of the body, prosthetic enhancements we might say.[McLuhan 1964]

[3] http://www.sci.utah.edu/~nmccurdy/Poemage/

[4] http://ovii.oerc.ox.ac.uk/PoemVis/

[5] The direct link to the recording is available here: http://media.sas.upenn.edu/pennsound/authors/Creeley/Harvard_10-27-66/Creeley-Robert_18_I-Know-a-Man_Harvard-10-27-66.mp3

[6] http://www.fon.hum.uva.nl/praat/

[7] https://wavesurfer-js.org/

[8] https://tla.mpi.nl/tools/tla-tools/elan/

[9] This is best experienced by visiting the URL for the Aligner and clicking "As" and "I". For the purposes of the printed page, please note that "As" is the first pitch curve in the line, and "I" is the second.

[10] I want to emphasize that while I am working to define the term *Machine-Aided Close Listening*, previous examples of its application exist, including MacArthur's and Clement's work. As we work to identify and define this term, other prior applications will likely surface, as well.

Works Cited

Artaud 1938 Artaud, Anonin. *Theatre and Its Double*. 1938.

Baraka 1966 Baraka, Amiri. "Black Art." *Liberator Magazine*, 1966.

Bernstein 1998 Bernstein, Charles, editor. *Close Listening: Poetry and the Performed Word*. Oxford University Press, 1998.

Brinkman 2016 Brinkman, Bartholomew. *Poetic Modernism in the Culture of Mass Print*, Johns Hopkins University Press, 2017.

Cabanne 2009 Cabanne, Pierre. *Dialogues with Marcel Duchamp*. Translated by Ron Padgett, Da Capo Press, 2009.

Clement 2012 Clement, Tanya E. "Distant Listening: On Data Visualisations and Noise in the Digital Humanities". *Text Tools for the Arts*, special issue of *Digital Studies / Le champ numérique*, vol. 3, no. 2, 2012.

Clement 2016 Clement, Tanya E. "Towards a Rationale of Audio-Text". *Digital Humanities Quarterly*, vol. 10, no. 2, 2016.

Clement and McLaughlin 2015 Clement, Tanya E. and Stephen McLaughlin. "Visualizing applause in the PennSound archive". *Jacket2*, 18 October 2015, http://jacket2.org/commentary/clement-mclaughlin-pennsound-applause. Accessed 9 April 2017.

Creely 2011 Creeley, Robert. "I Know a Man". PennSound, 2011. http://www.writing.upenn.edu/pennsound/x/Creeley/i_know_a_man.php.

Creely 2017 Creeley, Robert. "Machine-Aided Close Listening: Robert Creeley's 'I Know a Man'", edited by Chris Mustazza, PennSound, 2011. http://writing.upenn.edu/pennsound/poetry-lab/creeley/machine-aided.php.

Dworkin 2013 Dworkin, Craig. *No Medium*. MIT Press, 2013.

Ernst 2013 Ernst, Wolfgang. "Media Archaeography." *Digital Memory and the Archive*, edited by Jussi Parikka, University of Minnesota Press, 2013.

Evans 2012 Evans, Steve. "The phonotextual braid: First reflections and preliminary definitions".*Jacket2*, 25 March 2012, http://jacket2.org/commentary/phonotextual-braid. Accessed 7 April 2017.

Filreis 2009 PoemTalk Podcast. PoemTalk #16: "Because I am always talking: Robert Creeley, 'I Know a Man'," Hosted by Al Filreis, guests: Randall Couch, Jessica Lowenthal, Bob Perelman. April 8, 2009. https://jacket2.org/podcasts/because-i-am-always-talking-poemtalk-16.

Genette 1997 Genette, Gerard. *Paratexts: Thresholds of Interpretation* tr. Jane E. Lewin. Cambridge University Press 1997.

Johnson 1927 Johnson, James Weldon. *God's Trombones*. Viking Press, 1927.

Johnson 1935 Johnson, James Weldon. *The Speech Lab Recordings: Reading at Columbia University,* December 24, 1935. Ed. by Chris Mustazza. PennSound. http:// writing.upenn.edu/pennsound/x/Johnson-JW.php.

Kittler 1999 Kittler, Friedrich A. *Gramophone, Film, Typewriter*, tr. Geoffrey Winthrop-Young, Stanford University Press, 1999.

MacArthur 2016 MacArthur, Marit. "Monotony, the Churches of Poetry Reading, and Sound Studies". *PMLA*, vol. 121, no. 1, 2016.

McGann 1993 McGann, Jerome. *Black Riders: The Visible Language of Modernism*. Princeton University Press, 1993.

McLuhan 1964 McLuhan, Marshall. *Understanding Media: The Extensions of Man. McGraw Hill, 1964.*

Moretti 2013 Moretti, Franco. *Distant Reading*. Verso, 2013.

Mustazza 2016 Mustazza, Chris. "Decrypting the archive: Scholars and archivists gather to preserve and grant access to poetry recordings". *Transatlantica*, 2016. https://transatlantica.revues.org/8185.

Olson 2009 Olson, Charles. "Projective Verse". *Poetry Foundation*, 2009. http://www.poetryfoundation.org/learning/essay/237880.

PennSound n.d. "Williams, William Carlos". PennSound Author Page. *PennSound*. http://writing.upenn.edu/pennsound/x/Williams-WC.php

Spoerhase 2017 Spoerhase, Carlos. "Beyond the Book?". *New Left Review*, Vol. 103, January/February, 2017.