Plotting Poetry
https://doi.org/10.7592/PP2025.01.mustazza
Popular Voices: Computational Analysis of Poetry and Song. ELM Scholarly Press: Tartu 2025.

The Voices We Do: Surplus Inscriptions in the Poetry Audio Archive

Chris Mustazza

University of Pennsylvania, USA
mustazza@sas.upenn.edu
0009-0002-3076-037X

The following is a keynote lecture I gave for the fifth edition of the Plotting Poetry Conference, held in Tartu, Estonia on July 5, 2022.[1] It is meant to be heard, rather than read, and I hope you will do your best to imagine it as speech rather than writing. The ideas set forth in this piece are heavily entwined with the voice as medium. As such, I have consciously avoided converting this talk into a more standard academic article.[2]

Disclaimer & Introduction

I want to start with a disclaimer, a sort of content warning about this talk. In this lecture, you will find no thesis. I have no argument, in fact, I rarely have an argument—I'm a very agreeable guy. You will find no solutions in this talk, only problems. You will not come away from this talk knowing how to do anything or with a simple message you can pithily describe to others.

It's even arguable whether I, that is the person speaking to you, am using my own voice. I mean, what does it mean to own a voice anyway? Aren't I an amalgam of all of my life's experiences, and if so, why should my voice be any different? In my voice, you might hear some resonances and regional speech sounds of my parents' New York of the 1950s and 1960s filtered through my own time in the city in the 1980s. You might hear the influence of my teachers, my mentors, those I consider to be models of giving academic lectures. You could hear the genre of the video conference, of Microsoft Teams, as it were. The kinds of lilts and pauses endemic to this medium, uncertain of whether you can hear me or if my mic is even on (is my mic on? Can you hear me?, spoken with a hint of anxiety that you cannot, in fact hear me, as these words are floating into the ephemeral void). I guess I contain multitudes, or there are multitudes of Is, either way is fine for us.

[1] Editorial note: This article has not undergone peer review; it is based on the author's plenary lecture and is presented in its original form.
[2] A recording of the lecture is available here: https://youtu.be/iV9INJd0Tlo?si=bkkyFZR_stKGGr9o.

In any event, my goal today will be to present to you a problem and an opportunity devoid of any conclusion. The topic is simple: we have greater access than ever before to recordings of poets performing their own work.[3] But what exactly is recorded in this audio? Or put another way, who exactly is recorded? If the I speaking to you right now, that is the entity you recognise as Chris Mustazza, is plural and myriad, then why should the poets we hear be in these recordings be any different? If anything, considering the voices they do as part of their performances is even more complicated than parsing everyday speech.

Let me back up for a minute. Those of you who have been attending the conference for a while may have seen me speak about my work to locate instances of the sermon in poetry recordings. That talk, and this one, are premised on the idea that modernist poets often structured the soundscapes of their poems after what I term *apposite sonic genres*. We are so used to thinking about poetry as writing and classifying it as such, that scholars often neglect how sound recordings of poets might offer alternative classifications. In my previous papers, I looked and how the great poet James Weldon Johnson embodied black preachers of the early twentieth century in his poems.[4] Of course, Johnson is doing so explicitly, suggesting his connection to the Rev. A.W. Nix and preachers like him in his statement of poetics.[5] But what about other performances where the sonic reference is not given so plainly? We can definitely use our ears to say, "this performance sounds like [...] an ad, a political radio speech, a comedy monologue, etc." For example, we can hear the form of stand-up comedy in Louise Bennett's devastating anti-colonial performances,[6] and we can hear the sounds of Italian Fascism in the speech cadences of F.T. Marinetti's Italian Futurist manifestos.[7] That is, in all of these cases, the poets are doing voices, namely voices shaped on genres recognisable to an audience.

In my project, I worked to see if machine listening and artificial intelligence could locate these genres across a heterogeneous poetry audio archive, using what Tanya E. Clement has termed distant listening.[8] Could we, for example, train an algorithm to recognise what the sermonic sounds like (or I should say one particular strain of the sermonic) and locate instances where a poet is doing a sermonic voice across the entire 6,000 hours of audio within PennSound?

The process of doing so involved training a neural network using a corpus of early twentieth century recordings of performances that identify themselves as sermons. Surely, if the labels of the records said that they were sermons, they must be fixed reference point which we can use to train the machine, right? Or could it be that the preachers recorded on these records were themselves doing

[3] See, for example: PennSound, SpokenWeb, Les Archives Sonores de la Poésie, Voices of Italian Poets, etc.

[4] Mustazza, Chris. "In Search of the Sermonic: Machine Listening and Poetic Sonic Genre." *Computational Stylistics in Poetry, Prose, and Drama* ed. Bories, Plecháč, Fabo, DeGruyter, 2022, pp. 87-98.

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

[6] Bennett, Louise. "No Lickle Twang." https://www.youtube.com/watch?v=sW9GeQF-1bU.

[7] Marinetti, F.T. *PennSound Audio Collection.* https://writing.upenn.edu/pennsound/x/Marinetti.php.

[8] Clement, Tanya. HiPSTAS Project. https://hipstas.org/.

different voices, and if so, how can we train the algorithm to locate the one that we want?

You see the problematic that is taking shape. In training a neural network with examples of speech, we cannot know what facets of the speech the network is matching on and whether those are aspects of a sermon or … something else … regional speech patterns, for example, or even a voicing that is shaped by the preacher attempting to project his voice to the back of the church. The same considerations apply if we attempted to write a supervised learning algorithm where we would pick out what, to our ears, makes these recordings sermonic, describe it empirically, and write an algorithm to locate it. In both cases, we are dealing with a moving and morphing target.

So we started with a simple question, what and whose voices are recorded in any given performance of poetry, and we've come to an ostensible impasse, where does one voice end and another begin? Given that I do not have an answer for you, I was hoping you could tell me during the Q&A. But for now, let's aimlessly wander around the notion of voices in literature and how we might think about its application to poetry audio and everyday speech.

Voice(s)

What do you think about when someone mentions "the voice" in a literary work? So often, this singular term, is used to describe something about the written style of a work. One could imagine an MFA seminar where the topic is finding your voice. I wasn't sure what it meant, so I did what any good scholar would do and Googled it. Here's the authoritative word from Google: "In literature, 'voice' refers to the rhetorical mixture of vocabulary, tone, point of view, and syntax that makes phrases, sentences, and paragraphs flow in a particular manner."[9] Sentences and paragraphs? That's disappointing. Not utterance, phrasing, tempo, prosody, timbre, or any other dimension that would suggest actual speech, spoken language?

In short, voice is mostly used a metaphor. There is no actual voice in writing, only what we might reconstitute from printed characters in our minds, read through our own identities and imaginations. And, of course, some of the most important scholars of the twentieth century grappled with the nature of this metaphorical voice and its reception, including T. S. Eliot, Saussure, Bahktin, Derrida, and the list goes on. But I would argue that this is a technological problem, rather than a philosophical issue. Best to call your IT person rather than your ontologist or epistemologist for help here. These conversations are premised on the fact that print was perceived to be dominant form of literary dissemination for the entire twentieth century. The coeval rise of sound recording and access to it slowly eroded that reality, until today anyone can make their lived, spoken voices available using only the phone in their pocket.

In the era before easily accessible sound recording, poets worked to score the sounds of their free verse works to the printed page. Consider Charles Olson, who spoke about the typewriter as a device for the musical scoring of text: "For

[9] https://www.masterclass.com/articles/how-to-find-your-writing-voice.

the first time the poet has the stave and the bar a musician has had. For the first time he can, without the convention of rime and meter, record the listening he has done to his own speech and by that one act indicate how he would want any reader, silently or otherwise, to voice his work."[10] In other words, poetry exists first as sound, as the author's voice, and the printed page is a sort of ersatz sound recording of sorts. In the absence of an audible voice, it will have to be good enough and some performance instructions might help this sheet music to be interpreted. But it would be better if you could hear it with your own ears, directly from the poet.

All of this changes with the rise of literary audio, which picked up steam during the mid-century and only accelerated from there on. In fact, as a corollary to our topic of popular voices, I will mention that the first American poetry audio archives were made in linguistic speech labs, whose primary purposes were the study of dialects via human subject research.[11] While the archives created some of the earliest or only recordings of the most eminent modernist poets, like T.S. Eliot, James Weldon Johnson, and Gertrude Stein, their primary purpose was to record speech samples of everyday Americans.

So in the rise of literary audio, in contrast to the metaphorical voices in writing, we are presented with mediated representations of embodied voices. I wouldn't dare say "real voices", lest you take me to task in the Q&A, but maybe … realer voices? Could we agree on that? OK, how about … voices you can hear with your ears? Voices that cause your eardrums to vibrate as disturbances to air pressure? It's like the difference between sheet music and a concert, or the difference between libretto and an opera, or like … I'm trying to be literal here, none of this is a metaphor! Do you know what I mean?

But as we spoke about, and much to the chagrin of my empirical aspirations, every voice is plural. So what I would like to do for the duration of this talk is to try to enumerate some of the voices present in any one poetry recording, backgrounded performances of identity and social situation that I refer to here as *surplus inscriptions* to suggest that they exist beyond what many would consider part of the poem proper. But I would argue these murmuring, polyphonous voices come together to create the sonic form of the poem, spoken by a multitudinous author and received by an equally complex listener.

1 Personal and Literary Influence

It's no secret that we can come to inhabit the voices of those closest to us. I often ask my students: do you ever say something and then think to yourself "I sound just like my mother"? I don't mean the particular word choices, but rather the pronunciation, tempo, phrasing, or intonation of those words. As this concept of subconscious imitation relates to poetic performance, we can often hear aesthetic lineages performed as sound. Perhaps the most curious case of

[10] Olson, Charles. "Projective Verse." Poetry Foundation, https://www.poetryfoundation.org/articles/69406/projective-verse. Accessed Feb 9, 2023.

[11] Mustazza, Chris. "Speech Labs: Language Experiments, Early Poetry Audio Archives, and the Poetic Record." https://repository.upenn.edu/edissertations/3467/.

this phenomenon is between William Butler Yeats and Ezra Pound. A young Pound was Yeats' secretary and poetic admirer. Of course, it makes sense that some of Pound's performance techniques might derive from Yeats. But … how much of an imitation is flattery or homage and how much is … appropriative (or just weird)?

Yeats is recognised as one of the greatest Irish poets, and Pound spent his formative years outside of Philadelphia and at the University of Pennsylvania, where I teach. He later became known as perhaps the quintessential American expatriate poet, residing between England and Italy, until his ignominious extradition to the US on the charge of treason.

Let's listen to an early recording from PennSound, Yeats performing his iconic "The Lake Isle of Innisfree", recorded in 1936.[12] The voice is sonorous, mellifluous, and wistful. It extends the content of the poem, which is about Yeats' memories of his childhood home. It is about memory and its relation to presence and lack. The performance does not sound modern even for its time. Its chanted religiosity connects it to an earlier age of performance. Most importantly, Yeats sounds, well Anglo-Irish. We can hear the regional speech sounds that derive from Yeats' childhood home in Sligo and his upbringing in England. One of the distinctive characteristics of his speech is the rhoticity with which he rolls his Rs.

It is indeed uncanny to listen to performances by Pound in relation, such as this recording of "Sestina: Altaforte" recorded at Harvard in 1939.[13] Now, I don't know if any of you have been to Philadelphia, but I have lived here for most of my life and I have never encountered anyone who sounds even remotely like Pound does here. I would say, though, that the resemblance to his Irish mentor's performance is striking. Pound affects a kind of pseudo-brogue, rolling his Rs, and chants in the style of previous generation's poets. Pound's performance style goes beyond homage, it goes beyond pastiche, it drifts into unconscious caricature. We can say without exaggeration that Pound is doing a Yeats impression, and a pretty good one at that. I think I like Pound's Yeats better than Yeats' Yeats, in fact.

What we are hearing here is a very specific form of literary influence transmitted through speech, namely between poets who spent time in physical proximity to each other. We can hear the same kinds of influence in so-called literary scenes, where groups of poets who lived near one another came to perform like each other.

We could also look at what we might call … a case of mistaken identity. The poet Robert Creeley cited William Carlos Williams as one of his primary influences. Unlike Pound's direct relationship with Yeats, Creeley had never actually heard Williams perform, yet he built his own signature performance style based on what he *imagined* Williams sounded like, based on William's writing. Here's a sample of Creeley performing his hit single "I Know a Man".

[12] Yeats, W.B. "The Lake Isle of Innisfree". *PennSound*. http://media.sas.upenn.edu/pennsound/authors/Yeats/Yeats-WB_Lake-Isle-of-Innisfree_10-28-36.MP3.

[13] Pound, Ezra. "Sestina: Altaforte." *PennSound*. http://media.sas.upenn.edu/pennsound/authors/Pound/1939/Pound-Ezra_01_Sestina-Altaforte_Harvard_1939.mp3.

Note how he pauses at the end of each line.[14] The story goes that Creeley, so influenced by William Carlos Williams and his writings on the variable foot and his syntax-rupturing line breaks, was sure that Williams must have paused at the end of every line: why lineate a poem as such if he didn't? Well, he didn't. Williams reads straight through the enjambment in this 1942 recording of his famous "The Red Wheelbarrow".[15] Creeley had never heard the sound recordings of WCW performing and was influenced by an imagination of a voice. Williams' voice. A metaphoric voice in writing. This is an impression of a style that never was, yet this attempt to copy made something new.

Much more could be said about topic of personal influence, but in the interest of time, let's leave it at this: one of the surplus inscriptions in the poetry audio archive is that of aesthetic and personal influence.

2 Real or Performed Locality

We just spoke about Pound appropriating Yeats' identity as his own. The notion of regionality and place is often foregrounded in modern poetry, again with varying levels of, dare we say the word, *authenticity*. This is to say that poets perform place through the use of regional accents that are sometimes aligned with their own conversational speech sounds and sometimes ... less so. Sometimes the voices being done are from ... somewhere else ... someone else, varyingly problematic impersonations.

One example of this was Robert Frost's association with rural New Englanders. Born in San Francisco, raised in the small city of Lawrence, Massachusetts, and educated at Dartmouth College, Frost's upbringing was not that of a working-class New England farmer, the identity most of his poems would embody. Let's listen to a sample of Frost performing "The Code", recorded in 1933 at Columbia, University.[16] As Frost tells the frame story of a migrant farm labourer recounting to his current employer how he almost killed his previous employer, the aural mise-en-scène is crafted through Frost's application of regional accent, foregrounding the localism of the worker and dampening his accent for his performance of a generic employer. In other words, it is clear that the urbane Frost identifies more with the rural worker than with the bourgeois employer. While there are no recordings of it, Frost's "Death of a Hired Man" suggests similar dynamics, and in the recording of his canonical "Mending Wall", we can hear Frost doing the voices of two New England farmers.[17] Frost does not seek to mock his appropriated regional accents, but are they his to perform? Performances of the imagined rural, local, and everyman suffuse poetry recordings, an artifice that yearns for an authenticity, the desire to represent another,

[14] Creeley, Robert. "I Know a Man". *PennSound*. http://media.sas.upenn.edu/pennsound/authors/Creeley/Harvard_10-27-66/Creeley-Robert_18_I-Know-a-Man_Harvard-10-27-66.mp3.

[15] Williams, Willam Carlos. "The Red Wheelbarrow". *PennSound*. http://media.sas.upenn.edu/pennsound/authors/Williams-WC/01_Columbia-Univ_01-09-42/Williams-WC_01_The-Red-Wheelbarrow_Columbia-Univ_01-09-42.mp3.

[16] Frost, Robert. "The Code". *PennSound*. https://media.sas.upenn.edu/pennsound/authors/Frost/Frost-Robert_06_351A_The-Code-Part-1-Take-1_Speech-Lab-Recordings_New-York_1933.mp3.

[17] Frost, Robert. "Mending Wall". *PennSound*. https://media.sas.upenn.edu/pennsound/authors/Frost/Frost-Robert_10_355A_Mending-Wall_Speech-Lab-Recordings_New-York_1933.mp3.

to embody a voice that is not one's own. Frost's own New England accent is a voiced identity in this poem, yes, but it is paired with an overdone version of itself. These voices of place are surplus inscriptions to layer on to personal influence and sonic genre.

3 Voices Shaped by the Processes of Mediation

Finally, another layer of sonic inscription is the way technology (or the lack thereof) influences the voices we do. The voicing of much contemporary poetry, which is subtle, vulnerable, and delicate, is a product of the microphone era, the era of so-called electrical recording. The microphone allowed for this understated voice to be captured as audio. Before the invention of the microphone, in the era of acoustical recording, we can hear voices shaped by the need to project the voice. Performers were either accustomed to only live performances without recording, where their voices had to reach the back of the church or theatre, or to very early recording technologies where the performer had to bellow into an acoustic horn in order to be audible. Indeed, the microphone did not appear in commercial recording studios until 1925 and began its dominance of shaping sound with the crooning of Frank Sinatra, which could be said to be as much the microphone as it was the man.

Let's consider the founder of Italian Futurism, F. T. Marinetti, here. Marinetti is most famous for his Futurist Manifesto, which was published on the front page of Le Figaro in 1909. He called for a style of art that rejected everything from the past and perpetually attempted to remake the world anew through the latest technologies. It was a problematic movement to say the least, but it was influential on most strains of art, from poetry through to cooking and fashion. Marinetti was a bombastic figure, which we can hear in this recording of "Definizione di Futurismo" from 1924.[18] The defining facet of this work is Marinetti's characteristically aggressive delivery of his rules for the remaking of art. But it's more than aggressive. It's shaped, I argue, by Italian political speeches of the time, which were often given live in a piazza and broadcast over the radio across the country and into colonised regions in the north of Africa. It's unclear if they are electrical or acoustic recordings, but even if Marinetti had to project a little for the acoustic horn, his speech is modelled on one of public address. It's as if he is on a balcony bellowing at his aesthetic adherents below. It's as if he imagines this sonic assault to be cast out onto the airwaves. The imagination of the technology (or its lack) makes the voice.

Once the microphone became standard equipment, for both recording and amplification, its softened sounds became the norm and shaped performances. The deep interiority of the so-called Confessional Poets of the 1950s and 1960s was extended and formed through the sounds of the microphone. And performers who continued a theatrical performance style shaped around live spaces did so with the knowledge that they were subverting the current standards.

[18] Marinetti, F. T. "Definizione di Futurismo". *PennSound*. http://media.sas.upenn.edu/pennsound/authors/Marinetti/Marinetti-Filippo_Definizione-Di-Futurismo_1924.mp3.

Amiri Baraka's poems,[19] those of Charles Bernstein,[20] and Anne Waldman[21] all call attention to themselves by refusing the sonorous qualities enabled by the microphone.

Marinetti was blending a sonic genre, the political radio speech, and the sounds of the technology of reproduction to create his sound. Consider, here, the contemporary concept of Podcast Voice, which is the dominant voicing of podcasts shaped by both the conventions of the medium and the specific condenser microphones used in the production of the episodes. Turns out that we're not just doing voices, we're doing several at the same time.

In Conclusion

We started from the question of how we can empirically describe voices to identify via deep learning methods, and the philosophical question of whose voices were in any given vocal recording. Recordings of Pound contain his impression of Yeats. Recordings of Creeley, an imagined mentor. Recordings of Frost, performances of regional and socioeconomic differences. Recordings of Marinetti, a political agenda and a medium for disseminating it. How can we identify what voice is being done if multiple voices are being done at the same time, overlapping, intertwining, echoing, cancelling, and amplifying each other at various points? Or can a voice be separated, described with prosodic vocabularies, and algorithmically identified?

Here is the answer you've all been waiting for: Well, beats me! I don't know. When I started this project, I just wanted to find recordings that sounded like sermons. But now I'm not sure what it is—or who it is—the machine is finding when it clusters 'similar' recordings. But at the same time, I think the problems with this method are more interesting than any solution it could ever provide. The problems go beyond poetry and speak to everyday speech, to the deep heart's core (to borrow from Yeats) of identity and its malleability. The question is who are you? That's a plural you. Whose voices are we doing at any given moment? Where did we get them? What makes them us?

It calls into question the very notion of authenticity or a singular self. We *perform* our identities all day everyday, as Erving Goffman argued.[22] So why should performance or artifice be seen as something artificial or insincere? It would be insincere to suggest that there is a singular self outside of performance. The only question then is what the nature of this performance is.

[19] Baraka, Amiri. PennSound Author Page. https://writing.upenn.edu/pennsound/x/Baraka.php.
[20] Bernstein, Charles. PennSound Author Page. https://writing.upenn.edu/pennsound/x/Bernstein.php.
[21] Waldman, Anne. PennSound Author Page. https://writing.upenn.edu/pennsound/x/Waldman.php.
[22] Goffman, Erving. *The Presentation of Self in Everyday Life*. Anchor Press. 1959.