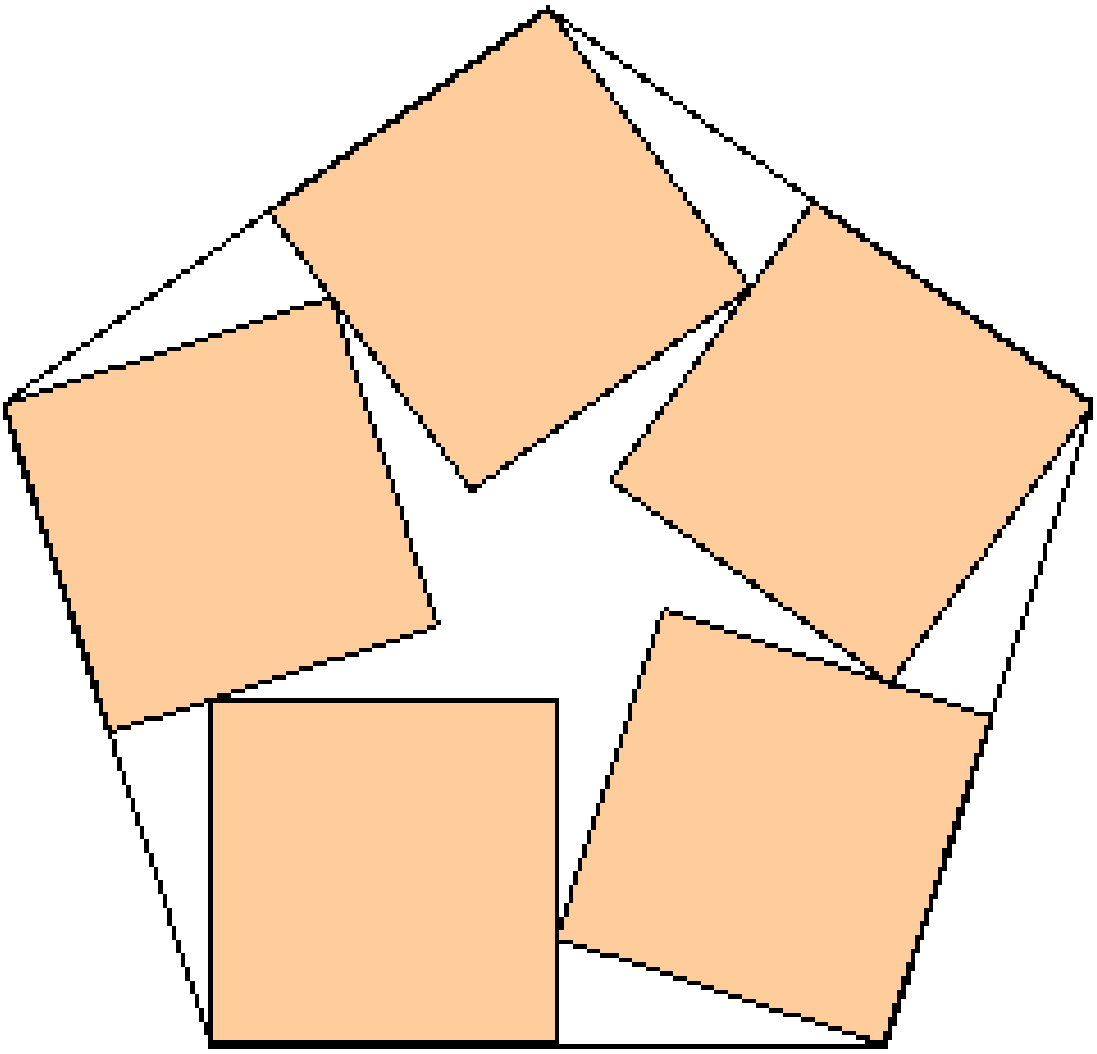


Poetics Of The Lacking Voice

Lawrence Mc Guire

master's thesis — Institute of Sonology at the Royal Conservatoire The Hague, 2026

Abstract

This thesis concerns the voice as a compositional material, specifically how presence can be generated through repetition and small variation rather than through combinations of fixed meaningful units. Taking Gertrude Stein's concept of insistence as a central reference, it develops theoretical and technical tools for working with vocal sound that sits outside or at the edge of language with the specific aim to liberate voice from its physiological constraints and sound expressions. Drawing on literary practice, phonetics, and articulatory synthesis, these tools are tested in a composition for natural and synthetic voice.

Acknowledgements

I wish to express my gratitude to my mentors Gabriel Paiuk and Ji Youn Kang for the time, effort, and thinking they have generously contributed and for the music they have introduced me to. Peter Pabon has still been aiding me from the sidelines by putting me into contact with voice researchers and generally acting as a friend, for which I want to thank him greatly. Regarding my music, probably the most influential person I have met in the past years is my collaborator Kieran Daly, who made me discover the work of Gertrude Stein and David Melnick and instilled in me hope for my own future in sound and text. The same can be said for Robert Francis (RM Francis), whom I would like to thank for his time, inspiring music and his enthusiasm. I would like to thank Raviv Ganchrow for the sporadic yet insightful talks and exchanges. Thanks to Jef Lambrechts and Cis De Gendt for allowing me to present my work over the years and for the ongoing amicable and musical companionship. Further gratitude and admiration to those no longer with us, but whose work lives on: Greta Monach, René Van Der Voort, François Dufrêne, Chris Mann. I would like to thank my partner Jiyoung for the care and love of that of an angel. Thanks Jacob, Roc, Yelim, Gene.

Contents

List of Figures

List of Tables

List of Sound Excerpts

Listen to the following sound excerpts through this link:

`https://hogobogobogoo.bandcamp.com/album/vox-sound-excerpts-ii`

1 *Crirhythme pour Bob Cobbing* (1970) by François Dufrêne

2 *Men In Aida (hum, vox, vox)* (2025) by Lawrence Mc Guire (as part of a forthcoming audio-publication on the *Iv (In vitro*)* imprint)

3 Timbral variations on aspirated [h] and vowel [e] by Gertrude Stein in her reading of her own poem *If I Told Him: A Completed Portrait of Picasso* (1923) 2m16–2m40

4 Recital of José Luis Castillejo's book "TLALAATALA," released on Algha Marghen (C alga11) with reading by Fernando Millán (2001)

5 *Simplex (Automatergon 72-24B var. 1 + 72-25A var. 1)* based on *Automatergon 72-24B var 1 + 72-25A var 1 by Greta Monach*, by Lawrence Mc Guire (2025) (as part of a forthcoming audio-publication on the *Iv (In vitro*)* imprint)

6 10 iterations of insistent shape modulations on a 3 second long NDVO insistunit.

7 An NDVO insistunit with no insistent modulations; this is the *global section*, synthesized in `VIS`.

8 An NDVO insistunit with 7 insistent modulations or $ALSs$, synthesized in `VIS`.

9 An NDVO insistunit with 20 insistent modulations or $ALSs$, synthesized in `VIS`.

10 Robert Ashley - *She Was A Visitor* from *Automatic Writing* (1967) 0m00-0m01

11 NKOAPP resynthesis; 15x iterations of Robert Ashley - *She Was A Visitor* from *Automatic Writing* (1967) 0m00-0m01

12 Synthetic articulatory morphing between excerpt 10 ($Vprim$) and a male shouting "Revolutionary!" ($Vsec$).

13 Unaltered excerpt from Chris Mann *goes a little 2fin* from `https://theuse.info/#`.

14 == 9.

15 Linearly increasing morphing from 13 to 14.

16 Logarithmic version of 15.

17 ALS^1_{LD} and ALS^1_{VO} for the insistunit "*o*"; used in Act I, phrase 3, first insistunit. *Four Saints in Three Acts* (2026) by Lawrence Mc Guire

18 *Four Saints in Three Acts* (2026) by Lawrence Mc Guire

Chapter 1

Introduction

> Before speaking open your mouth
>
> *Walter Marchetti*

This project trailed on from my interest in how text in its broadest sense finds ways to the mouth, or what paths the preverbal can be guided and transformed through to eventually arrive at some enunciation, by synthetic or acoustic means. These materials are what you could call "pre-vocal objects" on which I, the the seeer-reader-sounder, could either directly vocalize or, after intermittent source-changes (i.e., structural or sound transformations) vocalize into something else. The moment that crystallized my current modus operandi occured during a trip to New York. I visited various archives containing paintings, visual poems, verbal scores of Fluxus-adjacent artists, Concrete Poets, some of which I reinterpreted for a recital-performance on the same trip. What became clear through that experience is something I have more or less held to since: that text is my point of departure and that the audio-scriptic material within in it, be it phonetical, prosodical or otherwise, is to be addressed through a vocal engagement.

The interest I deal with here cuts across written, spoken, and heard language: phonetic systems, prosodic structures, nonsense voicings, the pre-articulatory and the post-articulatory. Because this project moves between languages and modes—English prose, symbolic notation, acoustic gesture, musical form—and because I use written English here as a kind of metalanguage for experiences conducted in other registers entirely, I want to be precise about what I take certain words to mean. What follows is a set of working clarifications, offered in the spirit of Dutch-German mathematician Hans Freudenthal's methodological prolegomenon to "Lincos: Design of a Language for Cosmic Intercourse" where he pauses before his project begins to clarify

his terms by stating axioms. (Freudenthal 1960, 1-45) Freudenthal's central concern was to design a language legible to an intelligent receiver (i.e., alien) with no prior knowledge of any natural language or its syntactic structures. What compelled me in his introduction was the methodological rigour of his preliminary distinctions, and the way those distinctions illuminate key elements of this project.

(i) on natural language and artificial language — Following Freudenthal (§0 03.), I use natural language in opposition to artificial language, where artificial means consciously constructed rather than evolved. Esperanto is his example; logistical notation (Peano, Russell, Whitehead) is another. My project works within and against natural language by constructing what I call mini-languages: temporary, local, formally constrained systems of vocal gesture that operate beside natural language. These mini-languages are artificial in Freudenthal's sense.

(ii) on the difference between a code and a language — Freudenthal strictly delimitates the distinction between a code and a language (§0 04.). A code is a system of transformation rules applied to an existing language and parasitises a host. Coding and decoding, he argues, do not presuppose understanding. They require only the mechanical application of substitution rules. Translation, by contrast, requires understaining.[1] When I work with vocal material that appears to encode or transform existing text, I am working without substitution rules and without a stable original to return to. The vocal forms I aim to produce in this project have no plain-language key. They are, in Freudenthal's terms, closer to a language than to a code, except that they are temporary, and accumulate no stable grammar, and are exhausted by the performance that generates them.

(iii) *formalist semantics* and what I do not want to do — Freudenthal is consistently critical of what he calls *formalist semantics*, the programme associated with Frege, Carnap, and Russell, which treats meaning as a fixed function from linguistic expressions to real entities, independent of context. I relate to his insistence that meaning is context-dependent and that there are no general rules of meaningfulness in any sufficiently rich language and that the effort to detach meaning from context is only possible if one is not primarily treating

[1] Freudenthal continues on the difference between coding-decoding and translation by noting that the rules of relation between spoken and written language are far more complicated than those of any cryptography. He writes that such rules "can hardly be handled by people who do not understand the language in question," even in languages with nearly phonetic spelling such as Italian and Dutch. A coding machine, he observes, can be much simpler than a machine for writing down or reading spoken language. He also points out that written language has grown considerably independent of its spoken origin, to the point where spoken and written English might be treated as two separate languages, while coded English remains entirely dependent on its plain substratum. He would allow the term "coded language" for coded English, but not the term language alone.

language as a means of communication. In sharing his scepticism, the presence of voice I indentify with does not map sounds onto fixed meanings or build a vocabulary with stable denotations (or referential in the sense that formalist semantics assumes). The sounds and voicings I produce mean in the way Freudenthal's dialogues on the word "to mean" suggest (§0 27): situationally, in relation to what surrounds them. Their meaning is bound to the moment and duration of utterance, the moment one engages the vocal artifice. It follows that the voiced objects in this thesis should be distinguished by whether they carry any denotational commitment at all. I introduce the terms denotational vocal object (DVO) and non-denotational vocal object (NDVO) to mark this distinction: as a continuum along which vocal material can move, sometimes within a single phrase, sometimes within a single phone.

(iv) speech and high-level music — I hold as a working axiom throughout my writings that speech is an instance of music at a high-level of complexity: the phonological, prosodic, and rhythmic organisation of the spoken word is a subset of a broader set of acoustic possibilities, of which music is the general case and speech the specialised one. The movement between speech and non-speech vocal sound is a movement within a challenging yet navigatable continuous acoustic space. Freudenthal's mathematical language made for the extragalactic layperson (§§0 06-0 09) offers a partial analogy. Because mathematical notation constitutes a genuine language[2] due to its syntactic structure being built on punctuation by brackets and the principle of freely interchangeable variables, it is sufficiently distinct from natural language to constitute something independent. Its meaning operates through internal structural relations. I find something analogous in musical form and in the vocal gestures that sit at the boundary between speech and music, especially that meaningfullness can be structural and pass as referential-less.

(v) translation and repetition — The word translation in relation to this project means something closer to what Freudenthal calls *quasi-general definition* (§§0 21-0 22): a demonstration by sufficiently many instances, from which a pattern can be inferred, without that pattern being formally stated, saying that my interaction with text or other to be voiced material is not subject to any translation by rule. Instead of using some code, repeated withhdrawal to the spoken is made, until a vocal form nucleates that carries some of the material's structural or sonic properties into a new sounding register. The main theme of this thesis is around repetition, or at least a specific kind of repetition called *insistence* which is central to facilitating this process. Each iteration of a vocal gesture differs from

[2] Not because it refers the way natural languages do.

the last by some small deviation of timing, breath pressure, or phonemic colouring. Trying to retain something like a prosodic memory across iterations: the rhythmic, intonational, and phonemic character of the original, redistributed through successive vocal acts.

(vi) constructing vocal presence — Vocal presence, as I understand it and how I bring it forth in this project, is created through iterative and formally constrained operations on DVOs and NDVOs. As psychoanalyst Denis Vasse puts it: "The voice is never represented: it represents, it is the act of presence which represents itself"[3]. (Vasse 1999, 106) I rely on psychic accumulation across iterations and their deviations rather than residing in any single utterance. Freudenthal assumes a receiver who understands his language by being able to operate on it (§0 12). I try to propose a related assumption about the listener: that what they receive is a vocal presence built through such operations. Whether that presence coheres locally, globally, or only asymptotically is the central question my thesis tackles.

More generally, this thesis describes how a vocal presence can be built from absence, repetition, and minimal deviation. Drawing on Gertrude Stein's concept of insistence which says that emphasis, and not repetition, animates language, engagements with my own voice and synthesized voice delineate how iterative operations on both DVOs and NDVOs can generate what I term a *presence-by-differential* ($\nabla^2 P$). In the first part, a literary-theoretical account triangulates Stein, Homer's "Iliad," and David Melnick's "Men in Aida" to argue that scripts can solicit vocalization without having to explicitly demand it. Second, Thomas Pynchon's concept of temporal bandwidth is used as a temporal model for insistence. Third, a formal, non-perceptual framework, borrowing from sheaf theory in mathematics models how local vocal identities cohere, but also fail to cohere, or even asymptotically approach a global presence, discussed across two works: José Luis Castillejo's "TLALAATALA," and Greta Monach's "Automaterga." Also, ideas Catherine Christer Hennix's around semantics in the Arts are brought up. Fourth, a technical account of the software environments `NKOAPP`, `Audio2Tract`, and `VIS` demonstrate how digital articulatory synthesis and Acoustic-to-Articulatory Inversion can simulate motoric independence in voice production, circumventing the coarticulatory constraints that fossilize natural speech.

These ideas converge in the text-sound composition *Four Saints in Three Acts* (2026), in which a natural and a synthetic voice share a single phonological stream, insisting on Stein's libretto

[3] Translated by me from French to English: "La voix nest jamais représentée : elle représente, elle est lacte dune présence qui se représente ou qui représente un objet pour un autre." This passage appears at the close of Vasse's account of a boy named Hector in analysis, tracing his path from a pre-symbolic immersion in the umbilical to him becoming a speaking subject, where the voice is presented as irreducible to representation: it is the act by which a presence represents itself, or represents an object to another. Vasse ties this to the concept of origin, arguing that the articulation of speech represses into the unthinkable the very voice that articulates it.

across three tiers of vocal granularity: phonic, morphic, and phrasal.

Chapter 2

Insistence

2.1 Eskimo Realities and Gertrude Stein

An interesting way of how words are used can be seen with the Eskimo people. The Eskimo hears the spoken word as a single event, where the word only becomes meaningful if it is put into action through the mouth. Media theorist and Canadian Arctic anthropologist Edmund Carpenter notes that "Eskimo isnt a nominal language; it doesnt name things which already exist, but brings things/action (nouns/verbs) into being as it goes along." (Carpenter et al. 1973, 39) The spoken word is to their ears an instance of reality that points to the order of nature, their beliefs and their existence at a specific moment and place. Carpenter brings up the example of the Inuit's naming a child at birth to reflect this idea:

> When the mother is in labor, an old woman stands around and says as many different eligible names as she can think of. The child comes out of the womb when its name is called. Thus the naming and the giving birth to the new thing are inextricably bound together. (Carpenter et al. 1973, 39)

Words are enunciated, temporally exist and then recede again in meaning. In Western thought, the birth of a baby would give rise to the name, but it is hard to imagine a spoken action partly bringing into being a being. English points to something, somewhere, sometime, while the Eskimo hunter conjures through the mouth, where "all words proclaim in themselves their own existence." (Carpenter et al. 1973, 38) It might be better to talk about form and function rather than a meaning-word duality:

> Words are like the knife of the carver: they free the idea, the thing, from the general

formlessness of the outside. (Carpenter et al. 1973, 43)

I would like to put this all into context with poetess and literary experimenter Gertrude Stein's stance on the use of words in writing and their sounds often take a more prominent role over sense.[1] Although she writes in a nominal language, being American English, she goes against it. Its nominal grammars are her subject of manipulation through reconfiguring, hacking, and re-aestheticizing the English literary logics. Similar to the Eskimo, language is understood as an act. To reject language as a system that points to given, stable objects and in doing so, form is brought into being through engagement in time. This 'form brought into being' is what I will describe in the coming paragraphs as 'making alive.' Stein does not feel settled with a system that relies on description and representational clarity. One has to stay with the language, and should not veer off into the tasty web of associations and definitions on which so many people's writing relies upon. On a speaking-tour in 1934 Gertrude Stein later collected in "Lectures in America," she vehemently talks about to what degree words are used in her writing. After going off on nouns, adjectives, slightly acknowledging adverbs and verbs, and then complimenting articles, prepositions and conjunctions, she mentions pronouns, seeing them "... not as bad as nouns" because they "can not have adjectives go with them." (Stein 1988, 214) What troubles Stein seemingly the most about nouns is their temporal inertia; their long-standing claim to things that appear already settled, already named. They are just not interesting to her just as nouns are not top of her list either. She even goes as far in saying that she suspects the adjective because it explains rather than presents meaning; finishes something prematurely; imposes judgment instead of allowing perception to happen. Adjectives tell you what something is rather than letting the thing it describes describe itself. If a nominal can be accompanied by an adjective the strength of its literary use to Stein diminishes strongly.

This concern becomes especially clear when Stein turns to proper names. She writes:

> Now actual given names of people are more lively than nouns which are the name of anything and I suppose that this is because after all the name is only given to that person when they are born, there is at least the element of choice even the element of change and anybody can be pretty well able to do what they like, they may be born Walter and become Hub, in such a way they are not like a noun. A noun has been the name of something for such a very long time. That is the reason that slang exists it is to change the nouns which have been names for so long. I say again. Verbs and adverbs and articles and conjunctions and prepositions are lively because these

[1] Gertrude Stein's work got introduced to me by my collaborator and guitarist-composer Kieran Daly.

> contain grammatical force and articulate relation, a transition, and dependence. As long as anything does something it keeps alive. (Stein 1988, 214)

She identifies a crucial distinction between naming as fixation and naming as event. Proper names, unlike nouns, are tied to a moment of creation where they are given. A name is attached to a life in motion rather than to an object stabilized across centuries. I find it useful to situate this point within the broader distinction between synchronic and diachronic approaches to language. A synchronic analysis presents a common noun as a stable, atemporal unit abstracted from the conditions of its use, but a diachronic perspective reveals that this stability is itself a sedimented history of accumulated acts of reference and meaning-making. William James, whose pragmatist philosophic teachings were formative to Stein, understood conceptual stabilization not as the suppression of process but as one of its outcomes: experience consolidates into habit, and habit into convention, without ever being severed from the flow of living that gave rise to it. Similar to what Deleuze and Guattari, in their "A Thousand Plateaus," develop as a pragmatics of language, where even the most apparently fixed linguistic constants are understood as the products of dynamic social and historical forces rather than as timeless abstractions. (Deleuze, Guattari and Deleuze 2009) What Stein inherits from James, I would argue, is the sensitivity to the way language conceals or reveals its own genesis: a common noun, for all the historical motion it has passed through, presents itself as self-evident, whereas a proper name retains the mark of a singular act in which *the being* and *the word* were bound together.

I see her favoring the idiom of the proper noun over the noun as an ontological preference. And in doing so, valuing linguistic forms that retain an element of contingency, choice, and change. In his "I Hear America," critic Vernon Loggins described Steins language as “thought in the nude—not thought dressed up in the clothes of time-worn rhetoric.” (Loggins 1937, 55)

In the following writings, I emphasize the enunciative act as the pivot around which all experiments and thoughts are gathered, where Stein’s stance and her idea that one has to do something to a word with the mouth will help me formulate a lot of things. This slight side-step makes a tangible link between Stein and the Eskimo’s practice of naming described by Carpenter above. The spoken name brings the child into being, where Stein remarks “. . . after all the name is only given to that person when they are born.” (Stein 1988, 214) Naming is an event that coincides with emergence. Just as a composer expects the player to play whatever is on the score page, so does Stein seemingly expect the reader to read with the mouth and not the mind, reverberating Italian composer and sound poet Walter Marchetti’s words in this chapter’s epigraph “before speaking use your mouth.”

The difference, again, is that Stein works within a nominal language—American English—and goes against its grain by choice. The Eskimo language structurally avoids nominal fixation and Stein reconfigures nominal grammar from the inside to what I would call *reviving the text*. This act of bringing into being is done through her singular use of repetition; or rather *insistence* on phrase, sentence and word where she claims to have learned its difference with repetition when living with her many aunts, who often said the same things but always in a different way each time. She spoke in the Lydia Mendelssohn theater on the subject "Portraits and Repetition" that “I am inclined to believe that there is no such thing as repetition,” she said. “The inevitable seeming repetition in human expression is not repetition, but insistence.” (Stein 1988, 174) At each iteration on the returning clause or word, the degree of emphasis should vary; otherwise repetition is doing and saying the same thing as in nothing other than taking space. The emphasis brings difference and occurs through varying stress and relational positioning. Psychiatrist and poet R.D. Laing in his book "Knots" describes the patterns emerging from his syntactical recursions as “knots, tangles, fankles, impasses, disjunctions, whirligogs, binds.” (Laing 1978, 2) What is definitely never done is adding content to make a difference, and is especially noticeable in the deliberate omission of the adjective. A useful sidestep is to consider her stance towards repetition as a stance of someone making music. Repetition or return is what I deem synonymic to musical form: the structural principle by which a theme, once stated, is departed from and brought back, transformed by everything that has intervened.[2] Stein’s insistence works with the same logic, but instead of the musical theme, the clause comes back, differently but remains what Laing describes as “. . . perhaps, strangely, familiar.”. (Laing 1978, 2)

These remarks are a response to a common criticism that her literary portraits are full of repetition, and are always talking about the same theme. It is not repetition, Stein claims, as long as there is emphasis. Coming back to the noun: an object already doomed by knowing what it is before even reading it; why then use something that is dead[3] due to an overdetermination? A rose, before any engagement, is already archetypically rose—botanical, symbolic, poetic, olfactoric. As long as anything does something it keeps alive, Stein would say. I claim insistence does this to these dead words, especially in Stein’s work. It is not that she does not use nouns in any of her writing, on the contrary. I think she points to its habitual mode of presentation where replacing the noun or ornamenting it with other things does not fix the issue; one has to re-activate it.

[2] The recapitulation in the sonata form, the recurring *A section* in the rondeau, in the fugue there is a return of the *dux* in new voices...

[3] Dead, as in motionless.

Take the famous *rose* example from Stein's poem "Sacred Emily" where one starts to see the inner-ear taking prominence over the ocular in the reading of iterated structures:

> Rose is a rose is a rose is a rose. (Stein 1922, 178)

Each "rose" occupies a different position in time and is framed by a different syntactic and rhythmic environment. The first rose is a naming of the character Rose; the second a confirmation that Rose herself is a rose; the third as insistence; the fourth as saturation or perhaps exhaustion. Stress can fall differently each time—on rose, on is, on the cadence of the phrase itself. The reader encounters the rose anew four times while not having the referent rose lose its semantic link and have itself magically deleted from the readers cognition; its inherent semantic meaning does not change permanently. What brings aliveness, then? Imagine entering the same house from four different directions. The logic of the house does not change, one still goes from out to inside, in to out, but the imagined world around the concept of the house does change. Each entrance reorganizes perception—light, orientation, anticipation. The house thickens as an experience and you learn how to inhabit it. Stein insists the reader or speaker to stay with the word for a time and long enough for the semantic chains to loosen, and importantly, to be resolved sonically instead of trying referentially. What brings aliveness is the different experience of encountering the thing and multiply the modes of access to it. Then, only can the noun's internal differences surface, make it loose its given burden. One has to return the noun as it does something again and in doing something, it changes how we meet it. The Inuk's ethos on action to assert presence in "that the universe acquires form, 'existence,' only through man the revealer" and is done with Language as their primary tool. (Carpenter et al. 1973, 42) To make the natural world a human world that can be experienced by action of the mouth. Logics of insistence effect and activate the psyche through a three-part criss-crossing of memory, attention, and confusion.

The motivation behind Gertrude Stein's diagrammatic style is reminiscent of the Lettrist's attempt at dissolving the word into the expressive letter, or, as with Ultra-lettrist François Dufrêne, to push language beyond the word toward a non-denotational vocal field where sound and corporeal utterance take priority. Isidore Isou formulated this trajectory as a necessary historical law: when an art form is exhausted, i.e., when attempts fail at generating new forms in the same framework and when standards set in—then the artist has to go over to a *chiselling phase*, in which the fixed thing is deconstructed or atomized into its smallest expressive units or modules. This deconstruction is usually gradual, for example we start from Hugo to Mallarmé, Apollinaire to the Futurists, then from Dada to the Lettrists, and onward to poésie sonore. Stein operates within this chiselling logic, though her instrument is structural rather than phonemic. She holds

strong opinions on the elementary units of language—word, sentence, paragraph—and works on them as such. Through a careful return to objects or structures of sense (word, sentence, paragraph) in time, one reads and peels layer by layer until the point where sense is taken over by vocal auditory properties.

This insistence is a technique for deconstruction and requires time to unfold. The reader's time becomes a convoluted and swirly-like construct. One often loses track of it all together in syntactically flattened and diagrammatically constructed sentences. In her "Two Short Stories," this convolution is clearly read and heard:

> He never asks whether he is going or coming. He always hears that he is coming, he always hears that he is going. He once heard that he had had what he had and he had what he had and he would have what he had and that he might have what he did have and he said that that was what he was wearing when he heard what was said and he knew what he had and he said that he heard what he heard and that he was coming and he said he knew he had been going. He heard all who spoke when he was hearing all he knew. This is Walter. (Stein 1972, 260)

Important to note is that the deconstructed image fades, but then slips back in its denotational garderobe once an emphasis is absent for long enough. That is not to say, this oscillation is not interesting to the writer. Go in the opposite direction and start with the non-denotative object (i.e., the phonetic unit) and try to make this object present in the cognition of the listener. Insistence does this. This is the "amplic phase" in which basic elements, that by themselves in relation to each other do not imply or explain its importance, of a framework are reassembled to do something again; being a style or form. I am interested in how the basic elements of speech, or sounds perceived as vocal can through a series of minimal deviations, carefully spread over time and placed "neighboring" units (i.e., through saying the same thing in multiple ways) conjure a presence. I want to give the mouth sound presence, other than it being heard as "an odd vocal sound" wherein the doing (insisting) is *the making alive*. The meaning or "aliveness" that flickers into being through insistent use I see as a durational object of *psychic resonance*; it decays out away once not returning to it anymore. I want to follow Stein's drive towards an attention to the sound shape but from the opposite side. Using insistent modulation on non-denotational sound shapes from a non-descriptive voiced sound to a sense-percept. However paradoxical, the term *sense* should be regarded in this context as non-descriptive and not a word of meaning that is fixed. Dufrêne and Stein both show a confidence, despite their different materials, that the non-denotational unit can accumulate this sense-percept and I find this *open problem* that

follows to be one of the more generative problems this study opens:

> Whether this accumulative quality belongs exclusively to natural language structures, or whether it can be abstracted into a broader principle of composition applicable to vocal sound that has not yet resolved into the word.

Some thoughts of Swedish composer, painter, poet and mathematician Catherine Christer Hennix I find useful here. Working across drone music, mathematics, and semiotic philosophy, and drawing on her studies with La Monte Young as well as her sustained engagement with Ngrjuna's logic of negation, Hennix describes generative structures as 'semantics rich/syntax poor'[4]; her framework is developed within semiotics and modal logic; I am extending its application to vocal and phonetic practice, which is a move she does not herself make explicitly. Hennix proliferates that the rules of combination of the elements of these systems are minimal, and that these parts are heavily associative. The richness is produced through the sustained pressure of return: the same element is held again and again, slightly altered. Especially her drone works express this vision where tone insists rather than develops in any harmonic or teleological sense, and through that insistence something does accumulate in the listener that registers as non-fixating presence.

I find comparable organizing principle across most of Hennix's practice. Her engagement with the text-sound world, particularly through her earlier work with synthetic and electronically processed voice, such as in the text-sound composition *Still Life Q**(1969), applies the same logic to phonetic material. In these works, a vocal unit, whether a syllable, a breath configuration, or a synthesized approximation of speech, is held in recurrence long enough for the referent to crumble and for something else to take its place. Her involvements in the Fylkingen group in Stockholm, where text-sound composition was developed as a distinct practice by figures such as Sten Hanson and Lars-Gunnar Bodin, situates her within a tradition already committed to the idea that the voice, when submitted to insistence and electronic transformation, can generate these *sense-percepts* that exceed what the word alone could carry. I draw on this dimension of her work as much as on her drone compositions, because it is here the 'semantics rich, syntax poor' principle is applied to the voice as a physical and temporal instrument.

I see it being structurally analogous to what Stein does with the clause and what Dufrêne does with the consonontal in his *Cryrithme* series.[5] Non-referential units are placed in adjacency, returned to with slight deviation, and through that temporal pressure they begin to act on the perceiver from within duration itself, oriented inward rather than outward toward a referent.

[4] See `https://www.frieze.com/article/catherine-christer-hennix-2026-review`.
[5] Listen to the *Cryrithme pour Bob Cobbing* in sound excerpt 1.

Stein and Dufrêne remain tethered, however loosely, to the recognizable architecture of language, while Hennix proposes the same principle at a level of abstraction that could hold for a sustained frequency, a repeated gesture, or a vocal sound sitting just below the threshold of the phoneme. I want to be precise about this distinction: applying her framework to voice is an extension I am making, one that I believe her logic supports without her having argued for it directly.

What all three figures share is the conviction that insistence on the small and the syntactically poor constitutes a different mode of generativity, one that works through time opposed to an accumulation of semantic content.

2.2 Silent Reading Aloud: Triangulative Study of "Illiad," "Men In Aida," and "Patriarchal Poem"

I have argued that Stein resists description and therefore places the burden of understanding on active recurring engagement with the text through enunciation. I hold a similar ambition for my own practice: to find structures that solicit the reader's voice without issuing an instruction to speak. What interests me is whether insistent operations on language, of the kind Stein performs on phrase, sentence, and word, can reclaim a vocality that the conventions of print have largely suppressed, and whether that reclaimed voice is one less beholden to normative modes of utterance. To understand how Stein reclaims vocality within a print tradition that has largely suppressed it, it helps to triangulate her work loosely between two poles: the first is Homer's "Iliad," where sound and voice were structurally inseparable from the text's transmission, and David Melnick's "Men in Aida," where semantic reading collapses entirely and vocalization becomes the only path forward. Vocalization as pre-understood or given, recovered, and forcibly demanded show how written language can solicit the mouth. Emphasis, iterative structures, or giving words presence all point toward a foregrounding of suprasegmental techniques[6] in the reading act. This section aims to ratify Stein as a sound poetess by comparing her work with David Melnick's "Men In Aida" and how both writers make the reader exchange sense for sound. Eventually this study should make clear how these techniques are a means to get to a voiced sound. Indeed, aside from Stein's operas and plays, her works are not described by herself as "oral" works, yet, acute attention should be given to what the written might sound like. Instead of thinking along the lines of "let's read what we see here," the reader should ask "let's hear what we read here." This is important because she has a consciousness of the spoken morphemic, phonic, and rhythmic features in the novel format; she is arguably practicing a linguistic art

[6] Commonly referred to as *vocalics*, paralanguage or all non-lexical elements of communication by speech.

rather than a literary art, all the while stylistically singular in doing so. In the context of seeing her work as music, her texts could be labeled as "language compositions."[7] (Robson 1959, 10) I would like to imagine Stein was not solely noting down her words and leaving it at that. As with any kind of music, it needs a performance followed by a reception. This writing needs additional listening. First, a reading aloud of the text-composition, then one silently, then a second reading aloud; I imagine this gives the writing a fair chance to be heard and understood. Still, the reader remains alone. Silent reading is up next for discussion in Gertrude Stein's work in relation to reading aloud.

"In the European tradition since at least Augustine, silent reading has been connected with practices of contemplation and the metaphysical business of making meaning" classical studies scholar Sean Gurd mentions in an introductory note in David Melnick's book "Men In Aida."(Melnick, David J. 2014, 9) Staying silent is still doing something, where one still stays with the word semantically and intellectually, but does not stay with its sound, or how it is heard when voiced. A step back is taken to make distance from the text logic and complex associations, and possible embedded messages or thoughts the writer might want to express. Contemplation is defined in the Cambridge Dictionary as "to spend time considering a possible future action, or to consider one particular thing for a long time in a serious and quiet way"[8] More concretely, one does not stay with the word or sentence, one brings it to another place for further musing and then brings it back to then carry on. Stein does not allow the reader to go to this place. The closeness with the word happens in my ears through using one's voice. Seeing Stein's writing then as a score to be able to enter a place where decoding does not have to happen. Similarly to David Melnick's "Men In Aida," the text itself paves a path to be traversed with the mouth and not with the mind. One that implicitly forces, subtly coerces one's vocal apparatus to engage, more than it explicitly instructs the text to be read aloud.

It is interesting to note that nearly all written works in the Antiquities up to the age of humanism and the invention of type by Gutenberg are meant to be read aloud in a social setting. (Gavrilov 1997) Silent reading was sometimes done in the reading of letters or shorter texts and was regarded as unpopular. Writing, reading's companion, also involved vocalization which was then noted down by a scribe. (Schrader 2010) From the 1st and 2nd century onward in the Roman empire, an interest was shown in private, personal writing. McMahon links this

[7] Similar to Edgar Allen Poe's language composition in his famous poem "The Raven," where even the term "Nevermore," he says, is based on logic following the "unity of effect." The sounds in the vowels in particular, he writes, have more meaning than the definition of the word itself. (*The Philosophy of Composition* 2025)

[8] From Latin contemplat- surveyed, observed, from the verb contemplari, based on templum 'place for observation.' In the Dominican Order, contemplation is seen as a condition for action; Stein takes a different path to this condition to action by avoiding contemplation.

to a growing awareness of the presence of psychological interiority. This idea of psyche is a precondition for contemplation. Homer's epic poem "The Illiad" is a cultural archetype of this text-voice-body dependence to bring out the tragedy, heroics, and poetic and literary prowess latent in the Ancient Greek language. It functions both as a canonical oral-literary work and as a historical timepiece that sets in print and sound the height of Greek poetic technique and cultural self-assertion.[9] It was written in verse which required an orator to read it. There were professional performers of epic poems, or rhapsodists, who added personal interpretation to the text, embodying it and participating in the story, who knew to bring out its emotional phrasing by heart through the epos (i.e., epic meter). We can distinguish this kind of reading aloud from reading aloud in Stein by its reliance and importance on the imagined or heard phonics. How do they both operate under silent reading and under speaking out loud? The sentences in Stein compared to those in Homer have a very different intention behind them. With Homer, the sounds and meter are technologies to enhance mnemonic stability; the meter allows the poem to be remembered, transmitted, and performed across time. Formulaic phrases ('rosy-fingered dawn,' 'swift-footed Achilles') fit the metrical pattern and assist the rhapsodist in sustaining long narrative sequences, while with Stein, unlike Homeric meter, which enables the listener to forget the words and follow the story, Steins diagrammatized sentences incentify the attention back onto the movements within the word. Secondly, "The Illiad" is a work of literally epic proportions compared to scriptures in modern poetry[10]. It makes sense that Homer wants the reader to move through the text where the rhythm aids in this aspect. Stein is unsure herself whether the work she makes should be labeled as poetry or novel, but aside from categorical thinking, her work often prevents narrative advance, because there is no plot in the traditional sense that should be carried forward. So, there usually is no obvious causal sequence for rhythm to support, and notably, any temporal progression collapses into a continuous present. The sound ornaments sense with Homer, whereas Stein's structures let sound compete with it, and so can regain significance.

Is there an alternate reading of the "Illiad" that can ignore these rigid qualities embedded in the suprasegmental elements related to tradition, narrative, and cultural memory? How can another path to its voicing be made so that the sound becomes a different determinant in the perception of its enunciation? An example that clearly asks and ultimately requires the reader to read aloud through the voice of a gay man is David J. Melnick's "Men In Aida." David Melnick (b. 1938 - d. 2022) was an American avant-garde poet who is often linked to

[9] Composed orally by poets over generations, likely culminating in a single master poet known as Homer.
[10] The standard "Illiad" consists of 24 books with a total of 15,693 lines of verse

Charles Bernsteins L=A=N=G=U=A=G=E Poetry movement, also known as Language Poetry, where members use prose poetry and longer forms to explore language as a medium, often disjointing the voice of the writer to challenge ideas of authorship and meaning. "Men In Aida" is a homoerotic, homophonic translation of the "Illiad"—Melnick listened to the Ancient Greek text as if it were English, translating the sound rather than the sense and drawing out the modern language he heard embedded in the ancient. Melnicks homosexualized translation, or “phonetic transcription,” emphasized the strong presence of homoerotic culture in ancient Greece. So, there is some mediation in nonsense word choices while accounting for comparable phonetics and his overt non-compliance to the (i.e., epos or dactylic hexameter). "Men In Aida" is undeniably a queer work, which is obvious from the recurrence of the word gay. According to writer Jed Rasula and sound poet Steve McCaffery, the poem “uncovers a homosexual pandemic riotously lurking in the very sound shape of Homers Iliad.” Or, as Sean Reynolds says “Melnick puts his mouth on Homers; he erects Homer into an icon of a certain gay performativity,” collapsing historical distance through phonetic intimacy. (Melnick 2014, 16)

Reading, in the context of text-print, implies avenues to an understanding of plot or narrative; my experience with "Men in Aida" describes a struggle over whether a word is nonce or sensical. Philologist Sean Gurd rightly compares it in his introductory text to "Men In Aida" to “the same way that beginning readers of ancient Greek get down to work: slowly, painstakingly, with only the slightest hope of reading fluently, but with a quickening sense of the extraordinarily fine craftsmanship that only such slow reading can produce.” (Melnick, David J. 2014, 8) Due to this frustrating quality, I can only describe it as a form of intentional unresolvability in the silent reading of the text. A quality I wanted to amplify and test in a concert and audience setting. The performance was part of a suite and was essentially my take on a simultanous poem or *poéme simultanée*, popularized by the sound poetry group "The Four Horseman." Three voices recite the opening of Book I of "Men in Aida" simultaneously. One voice delivers the Ancient Greek in dactylic hexameter, taken from a recording by the classicist Stanley Lombardo. A synthetic voice resynthesizes the Greek recitation into a hum, produced each time a pitched segment is detected. A third voice oralizes Melnick’s text through my mouth, following the phonetic proximity and humor of his writing rather than any semantic fidelity to Homer. The Greek recitation carries metrical, historical, and prosodic structure. Heard alongside Melnick’s version, it loses its position as origin and becomes one layer among others. The hum sits between the two human voices without belonging to either, sort of like a pseudo-linguistic judge between them. A recording can be found in excerpt 2. I am faced with reading it aloud as the only way further down the verse: Gurd also notes he experienced a similar reaction as the only way in dealing with

the text. The text induces “us to give up meaning, to open our mouth and sensualize the reading process.” (Melnick, David J. 2014, 9) Melnick shows us that the only understanding present is a sonorous one, not semantic. The oralization is necessary in case semantic reading fails, and it is here that the vocable takes form. Under the guise of unresolvability, a transformation of urgency from text as sequence to text as sound occurs. His broken homophonies make sure that the text gets livened up, but somehow resisting a decoding of the heard material, giving all the more reason for it to be recited. On top of all this, mental conjurations of meaning arrive through sensations of recognition, absurd humor, erotic charge, and disturbance, but misses—in a positive manner—a narrative continuity.

The work of Melnick is unreadable, in the normative sense of reading, while Stein’s work is comparably readable. They both nudge the reader to traverse the opaque and convoluted language from the optical to the auditory in their own way: Melnick by a phonetic transcription that has a quality of unresolvedness, of quasi-words and fugazi expressions; Stein tricks the reader more subtly to trip over their own feet cognitively by participating in her iterative syntactic language games and confronting them with willful incoherence. Important to note is that little of her work became known or entered the American literary canon for a long time, and when it did, it was rarely recognized as innovative within the literary and verbal arts. Stein’s work “encountered resistance of mainstream culture to the new and unconventional.” (Stein 1997, v) For example, “Stein composed ‘The Making of Americans’ from 1903 to 1911, though it was not published until 1925, in an edition of 500. The novel was not reprinted in full until 1966 by Fluxus artist and poet Dick Higgins’s Something Else Press (New York)[11], making the book available to a new generation of writers and artists,” given in a statement on the organizer’s website. (Triple 2013) Sound poet and radio producer Charles Amirkhanian gave a series of talks and readings of Stein’s "Lectures in America" on his long-running radio program *Ode To Gravity* on public radio KPFA-FM. In another installment of the same radio program, he calls her work "transrational" (usually translated from the Russian Futurist’s *zaum*, meaning “beyond the mind” or “beyond reason”). (*The History of Sound Poetry* 1976) Amirkhanian insists her work operating past conventional semantic logic, yet always retains intentional and structured. Featured in the same show is a reading by Gertrude Stein herself in 1923 in Paris describing her good friend and Cubist painter Pablo Picasso. At the end there is a passage of insistence on the word *he*, which becomes more regarded as an objectified sound where variations on the timbre of aspirated [h] and vowel [e] as well as the stress positions are under investigation see

[11] From 1974 to 2000, Paula Cooper Gallery hosted marathon readings of the classic novel "The Making of Americans" around New Years Eve, including Higgins, Alison Knowles, among many others.

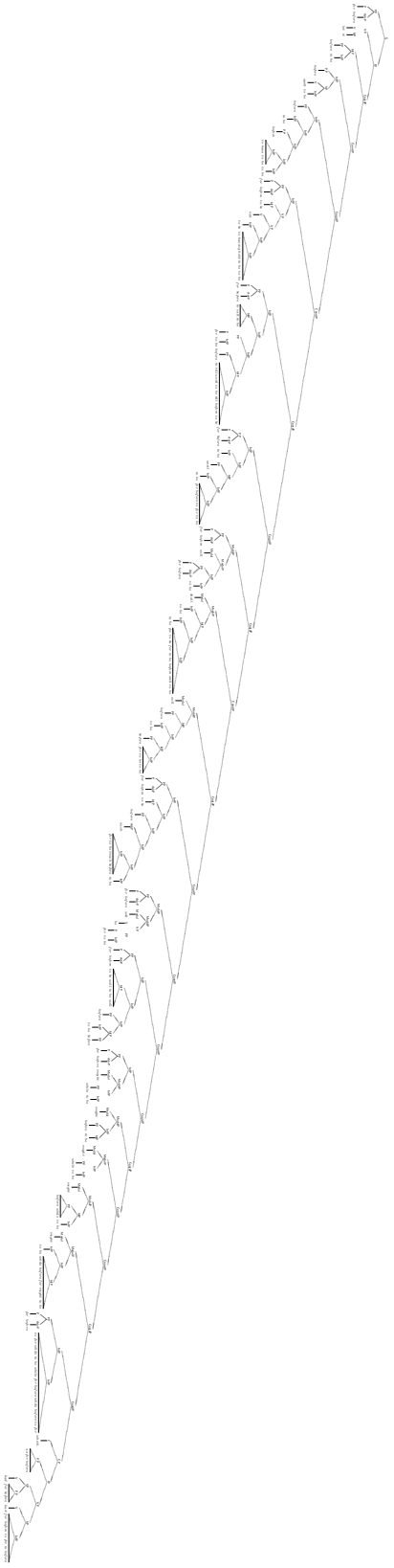
excerpt 3. There is much reason to consider Gertrude Stein as a sound poet, but her genius is realized as strongly in her written work. "Patriarchal Poetry" (1927) is exemplary in showcasing Stein's peak in phonetic and syntactic experimentation. A parse tree (or *concrete syntax tree*) of the opening sentence of "Patriarchal Poetry" is given in **??** to get a visual sense the hierarchy of how nested these structures can be. (Mc Guire 2026, 22-25) As one of her most prominent and agile readers, Marianne DeKoven contends that this "lengthy piece [. . .] not only defies interpretation, it defies reading." It shares a similar frustration found in the verse of "Men In Aida." Although Stein's approach or intention seems to be more involved with a music making in my opinion in terms of presences of pulse, rhythm and melody, writer Eric S. Neel mentions "To read this piece involves both a hermeneutic pursuit of meaning and a risky engagement with interpretive uncertainty. This risk is most profoundly figured in a readers willingness to read 'Patriarchal Poetry' aloud, for when we do, we generate an unpredictable and heterogeneous phonotext." (Neel 1999, 89) Neel follows this up by identifying what I see as the distinguishing factor from the silent reading demanded by Melnicks work,

> In "Reading Voices," Garrett Stewart persuasively argues that even silent reading involves a "somatic quotient" that involves the organs of vocal production, from diaphragm up through throat to tongue and palate (1-2). He suggests, then, that reading always already participates in the sounds of language, and that meaning echoes and resounds in words, and in the relationship between words, in order to trouble the semantic intent of the script. So, as I go to the trouble of suggesting that we read Patriarchal Poetry aloud, I am emphasizing an already present relationship with the text that persistently manifests Steins language as sound and music precisely insofar as these phenomena are the unpredictable product of an encounter between writer, text, and reader. (Neel 1999, 101)

The somatic quotient in "Men In Aida" is markedly different in both degree and function. Text in Stein solicits vocalization as a risky but generative extension of interpretation. With Melnick on the other hand, the phonetic transcription forecloses interpretation in advance and insists on "sensualization." Arguably, a semantic hypothesis is tested through sound with Stein; Melnick wants the reader to submit to the muscular labor of articulation itself. This distinction clarifies why Steins work, despite its resistance to semantic closure, remains comparatively readable for the literate person on the other side of the print. In the lecture of "Patriarchal Poetry," the text confronts the interpreter with its own oral agility or—borrowing from Walter Ong—*orality*. (Ong 2005) Then, by attempting to achieve uninterrupted performance of the text, the only

attention that is possible is an attention determined to conclude the long phrase. There is no space for semantic interpretation, because phonetic closure is of a higher importance; there is more at stake if one physically fails to complete the sentence than contemplatively. Insistency turns into persistence. The scholarship around Stein's feminist politics argues that play in sound and rhythm is a proposal to disarm a Western patriarchal logic. If read through this idea, the unreadability due to phonetic opaqueness mobilizes her to form an epistemic resistance. She makes a claim for a possible "reading through patriarchy" because she still employs phrases and expressions that echo patriarchal articulations.

Figure 2.1: Parse Tree for opening phrase of Gertrude Stein's prose-sound poem "Patriarchal Poetry"



Failing to read in the normative mode generates another logic of its own, seemingly generating an autophagy of the normative structure. Patriarchal language is used to digest its own authority. In line with Stein's philosophy, this anxiety can also be minimized instead of completely traversing and ultimately resolving it. To balance the weight from persistence back to an insistent quality involves making the text readable again on the condition of no vocal breakdowns, and, importantly, making sure the orator concurrently holds dearly on to an audio-scriptic attention. There is a flow missing in the poem where insistence barely catches on. A high chance exists that the reader encounters Stein's repetitive verbal structures as obstructive rather than having an automatic and mantric function. An able reader will struggle less. One way to allow ease of reading is to intervene in the typography. Bionic Reading® is a candidate for helping the struggling reader that supposedly aids the reading flow by bolding the first few letters of a word.[12] Applying a bionic highlighting to the following second sentence of the opening of "Patriarchal Poetry,"

> For before let it before to be before spell to be before to be before to have to be to be for before to be tell to be to having held to be to be for before to call to be for to be before to till until to be till before to be for before to be until to be for before to for to be for before will for before to be shall to be to be for to be for to be before still to be will before to be before for to be to be for before to be before such to be for to be much before to be for before will be for to be for before to be well to be well before to be before for before might while to be might before to be might while to be might before while to be might to be while before for might to be for before to for while to be while for be- fore while before to for which as for before had for before had for before to for to before.

turns the text into the following form by simply increasing font weight of letters:

> **F**or **bef**ore **l**et **i**t **bef**ore **t**o **b**e **bef**ore **sp**ell **t**o **b**e **bef**ore **t**o **b**e **bef**ore **t**o **h**ave **t**o **b**e **t**o **b**e **f**or **bef**ore **t**o **b**e **t**ell **t**o **b**e **t**o **hav**ing **he**ld **t**o **b**e **t**o **b**e **f**or **bef**ore **t**o **c**all **t**o **b**e **f**or **t**o **b**e **bef**ore **t**o **t**ill **un**til **t**o **b**e **t**ill **bef**ore **t**o **b**e **f**or **bef**ore **t**o **b**e **un**til **t**o **b**e **f**or **bef**ore **t**o **f**or **t**o **b**e **f**or **bef**ore **w**ill **f**or **bef**ore **t**o **b**e **sh**all **t**o **b**e **t**o **b**e **f**or **t**o **b**e **f**or **t**o **b**e **bef**ore **st**ill **t**o **b**e **w**ill **bef**ore **t**o **b**e **bef**ore **f**or **t**o **b**e **t**o **b**e **f**or **bef**ore **t**o **b**e **bef**ore **su**ch **t**o **b**e **f**or **t**o **b**e **mu**ch **bef**ore **t**o **b**e **f**or **bef**ore **w**ill **b**e **f**or **t**o

[12] When we fixate a part of a word—the fixation—the necessary information is passed to our brain so that we can match our long-term memory with the available representational products. The jerky phase between fixations is called a saccade. During the saccade jump, we are unable to take in read information and pass it on to our memory.

be **f**or **bef**ore **t**o **b**e **we**ll **t**o **b**e **we**ll **bef**ore **t**o **b**e **bef**ore **f**or **bef**ore **mi**ght **wh**ile **t**o **b**e **mi**ght **bef**ore **t**o **b**e **mi**ght **wh**ile **t**o **b**e **mi**ght **bef**ore **wh**ile **t**o **b**e **mi**ght **t**o **b**e **wh**ile **bef**ore **f**or **mi**ght **t**o **b**e **f**or **bef**ore **t**o **f**or **wh**ile **t**o **b**e **wh**ile **f**or **b**e-**f**ore **wh**ile **bef**ore **t**o **f**or **wh**ich **a**s **f**or **bef**ore **h**ad **f**or **bef**ore **h**ad **f**or **bef**ore **t**o **f**or **t**o **bef**ore.

I claim that this phonological alleviation of the oppressive structure—being English—does not defeat what expressive promises the feminist transrational reading brings to modes of subjectivity. When voiced, tension between sense and sound stays present and the audition to them keeps recalibrating one's cognition.

Despite their differences, Stein and Melnick both displace reading from the optical toward the auditory, compelling a voiced engagement with language. Melnick displaces it to the mouth by a making silent, interpretive reading untenable; Stein achieves it more subtly in some of her work through insistence, where phonetic obstructions keep her work formally readable yet cognitively unstable. In Stein, this instability operates as an epistemic and feminist resistance to patriarchal semantic closure, using sound to erode inherited linguistic authority from within. Read through rather than for meaning, her texts unfold as durational and bodily events, in which repetition intensifies presence, and thereby make a strong case that insistence can work as a temporal and perceptual force guiding the attention of the interpreter toward the sonics of what is in front of them.

2.3 Insistence and Time

Directing attention from the referent to the sound is one thing, as we may deduce from Homer, Stein, and Melnick in the previous section, but then directing attention to movements within a sound is another thing. I am looking to create conditions for intensive attention to morphological shape in voice, both in the sensical and non-sensical contexts. The temporal importance of insistence and how it can be considered a formal device to create these conditions is given some further time in this section. A theoretical account of what insistence does to time and perception in the writer-time and reader-time, as well as the voiced-time is given. Additionaly, the laws govern time and perception within denotative and non-denotative (i.e., sensical and non-sensical) voiced structures, respectively abbreviated as *DVO* and *NDVO* ("D" stands for denotational, "O" stands for object, and "V" for vocal). Later more on this. An auditory attention takes place that incentivizes a fixed attitude towards the pre-verbal and verbal are not of concern here: the attention that dwells in the thickness of the Now is of larger interest. Discerning a linear and discrete model of reader-time seems like a good start.

Insistence can resist memory retrieval for contemplative activities, as it functions as an anti-mnemonic device that keeps attention close to the present moment of the text rather than reaching backward toward an accumulated meaning. In written language this dynamic is quite particular because unlike voiced time, the text is fixed while the reader moves on, creating a structural displacement between writer-time and reader-time where insistence can either deepen or dissolve the reader's sense of Now. A model of this displacement is found in novelist Thomas Pynchon's concept of the temporal bandwidth Δt. Thomas Pynchon's application of this literary device Δt can expand or contract a character's, and a reader's presence over time. Interpreted loosely, the larger one's temporal bandwidth, the better one makes decisions in the Now while accounting for the perceived past and imaginary future. An alternative description for Δt is that it is a measure of how far one takes the sensed thing into the realm of thoughts. Two characters in Pynchon's "The Crying of Lot 49" are narrative experiments in this temporal bandwidth. (Pynchon 2006) For example, one of the story's characters, Mucho Maas, has a small bandwidth that makes him incapable of integrating experience in time and therefore causes Mucho to feel anxious, while another personage called Tyrone Slothrop spans multiple temporal registers, embodying more of a spectral state of mind. Slothrop's body and psyche register cause-and-effect chains across years, even before he himself understands them. Over the course of the story, both narrative presences eventually disperse into being all everywhere and nowhere ate the same time. At the contracted end the sensed thing never reaches thought, and at the expanded one,

Δt spans so far that presences dissolve altogether.

The Now is Stein's territory and her writing occupies the contracted end of this spectrum. Pynchon maps and composes temporal bandwidth through his character-building and not specifically through language itself, Stein contracts Δt through heavy repetition and diagramming of linguistic forms, so that reading itself becomes a present-tense event. Opposite Pynchon's narrative presence, an automatic, sonic presence creates itself. Under insistent action on the word, associations with a personal past, technological history, geopolitical or mythic time disappear with each emphasis on it. The bandwidth shortens to the degree where it collapses into a saturated point such that each following iteration is complete in itself. This word-sound encroaches on the semantic, and in it, builds up towards something semiotician Roland Barthes refers to as punctum in his final book "Camera Lucida: Reflections on Photography." (Barthes 1980, 94-97) I am taking the liberty to use the idea of the punctum to describe a non-descriptive detail in voice that arrests the listener and pushes away the subjectivity one might have with the voiced phrase.[13] Piggybacking on the compression of Δt, one gets the accumulated pressure of what came before in exchange. Something interesting happens with respect to what type of essence we are listening to. In a word or phrase, it is important to note that one must discern multiple essences, or quiddities, that can be simultaneously present at one point in time;

1. semantic quiddity (lexical meaning, conceptual role)
2. phonetic/sonic quiddity (its acoustic identity)
3. what it does in use as in pragmatic quiddity
4. how it appears as in typographical quiddity

Linguistic forms have competing claims to essence depending on which ontology is prioritized. Here, a push away from a semantic "whatness" is imposed on the reader to the extent that the word or phrase becomes an event that corresponds to contracting temporal bandwidth. More useful here is the Buddhist Sanskrit concept aligned with the term *tatht* or *suchness*. To what the word is actually subjected to by the reader when realizing that this quality has no conceptual presence, and without conceptual presence, the word cannot be retrieved from memory or projected into anticipation, essentially saying that it can only be encountered again, new each time. This is the experiential consequence of a fully contracted Δt where pure recurrence hinders meaning-building.

Natural language can be seen as a highly structured and complex system. Language becomes noisy in the oral transmission of the word. On paper, language is ordered and clear, but when

[13] Not to be confused with Barthes's *grain de la voix* or grain of the voice.

enunciated and interpreted, language becomes a victim of its own disposition. Speech is hereby labeled as a chaotic thing. To bring order to the chaos of associations within interpreted language, insistence can revert it back to order, both in the reading as in the voicing of text, where the interpretation still implies subjective categorization based on intensiveness (i.e., more and more of this until this is all there is) instead of the usually taxonomic kind (i.e., this versus that kind). It avoids having to ask the question “Is this noise or a pattern?” Take, for example, Oedipa Maas in "The Crying of Lot 49." Oedipa encounters a chaos of signs—the Tristero, the muted post horn, the acronyms—and tirelessly attempts to sort these into sense. But she pays a price. Her intepretative labor results in episodes of anxiety and produces paranoia because of the uncertainty about whether she is finding real patterns or imposing false ones. Temporal bandwidth is experimented with on the narrative and character development level in Pynchon’s novel, while the Δt of the reader is more or less contracted in Gertrude Stein’s work. The “this-goes-there” and “that-goes-here” mentality towards stabilization is opposed by a "this with the this" attitude in Stein’s prose: a stabilization that I suspect aids me in facilitating non-denotative vocal presences. The sense in the text is saturated through a play of linguistic forms to avoid hierarchies based on the separation of different things and a density of the same thing results in a logos. The word “rose” starts to lift up, gaining buoyancy, all because of saturating the perceptual field with more roses, similar to satiation. The varied return to the rose shifts from extensive (covering ground) to intensive (deepening contact) attention. An intensiveness that brings with it a counter-taxonomical stance that by itself avoids the question of “what belongs where?” And in doing so, this intensiveness clears the perceptual field for something else entirely. Where Pynchon’s characters are burdened or scattered by the weight of accumulated time, Stein’s insistence contracts bandwidth until only the present moment of the word remains. It is within this contracted Now that the conditions for intensive attention to morphological shape in voice are possible: both in the sensical and nonsensical.

2.4 Insistunits and Insistent Modulation

My interest lies in working with iterations of micro-deviations on vocal shapes where the emphasis becomes the content I want to compose with. Emphases that foreground a difference or a *delta* between a current and previously perceived voice sound shape. This requires further commentary on cognitive products resulting from insistent operations on DVOs and NDVOs and how shape-iterations associated with nominal and non-nominal structures differ from each other in terms of linguistic affect and sound sensation.

Emphases in DVOs are largely created through neighboring elements; creating the *syntagmatic* and morphic (word-like) emphasis in a denotational system (e.g., language) is only possible with a syntax present. Syntax in language can be loosely described as a set of conventions governing the order of linguistic forms and the ways in which those forms may be correctly related to one another, independently of whether those are morphologically or semantically well-formed. "I drinked a fine gueuze yesterday evening" is syntactically correct because the words are in the right order and relate to each other in the right way. It follows the strict subject-verb-object word order of Natural English (i.e., linguistic topology). It is grammatically incorrect because it is morphologically incorrect: the past tense morpheme *-ed* and the verb *drink* do not combine with each other in any of the standard dialects of English and so by having this taxonomy, humans can plan their speech ahead. While reading, this learned taxonomy aids in predicting and neurologically priming the most likely word candidates in a person's mental lexicon[14]. So, morphology deals with the internal composition of words and syntax deals with the combination of words.

Ostensibly, Stein deals with the composition of words by dealing with the combination of words, subtly diverting attention to their text-sound through a deep know-how of the English morpho-syntax. Stein never intervenes in the construction of morphemes; she insists on words, sentences, and phrases iteratively; the sonic impression of the word left behind prevails above semantic attention that natural language initially asks for. Within this rigid structure, the writer seeks techniques of nuance that in spoken word are givens upon verbalization, such as conveying the correct speech rate, the absence of a vocal identity or idiolect, pausal duration, and stress position. This interests this project on the aspect of how others—not only sound poets or vocalists at large—work in written language to inject hermeneutic ambiguity through some know-how of the linguistic and acoustic mechanics of sound over time.

Aside from insistence working as an intensive attractor of temporal density in the context of DVOs—seen in Stein's work—can insistence modulate temporal density, perceptual resolution, or even attentional bandwidth of heard NDVOs? And if so, do insistent techniques on NDVOs share the same quality of generating vocal presence through difference? That is, the emphases or insistent modulations on a non-denotational structure effect similar changes regarding constructing or deconstructing form in the aural. It might be good to discern what I mean by an NDVO and compare insistent repetition of these against, for example, instrumental sounds. First off, NDVOs should have the following time-phonology shape characteristics;

[14] psycholinguists define it as the mental collection of words and representations

1. no repetitive structures or "reduplications"
2. short duration
3. do not belong to any existing lexicon (i.e., learned lexical articulatory formations)

Repetitive phonic structures are often heard as onomatopeic and ideophonic. These seemingly nonsensical utterances carry an iconicity. Child-directed speech or baby talk such as the reduplication of -ba or -da is a clear exemplar. Along with a phonic likeness across syllables of ideophones, syllable times are often similar or even in length, meaning an NDVO should avoid having reduplicative qualities in both temporal and spectral dimensions to avoid sensory-imagistic associations.

What is perceived as familiar or unfamiliar is largely conditioned by the listeners exposure to particular phonetic inventories and phonotactics[15]. Most of the sounds coming from the mouth are often geographically and socially distributed, but often still allow for phonaesthemic or affective associations that can overlap different linguistic systems. An example of phonaesthemic similarity is the Austronesian phonaestheme ŋ which occurs in initial position in words related to 'nose' or 'mouth,' appearing in forms meaning 'chew, masticate, ruminate' such as Amis ŋafŋaf, Toba Batak ŋaltok, and Trukese ŋʈ—none of which are cognates—reflecting a cross-linguistic tendency to associate a given sound with a shared semantic domain across genetically related languages rather than through common ancestry[16]. (Bergen 2004, 2) An NDVO differs from the categorization of the phoneme in that it can consist of multiple perceptually distinct units of speech. All phonemes, meaning those with little to no ideophonic associations are NDVOs, but not all NDVO's are phonemes. NDVOs can be strings of phonemes or nonsense strings. The vocal phenotype described above does not imply that this sound need be naturally sourced: synthetic sounds also belong to this class as long as their mode of production can be inferred by the listener as voiced or as vocal. That is, as arising from an articulatory model of phonation, independent of biological origin. Aside from these possible associations, this proposed insistence is a temporal device that directs attention toward conferring a halo-like presence on an voiced object; with each iteration, this halo "deafens" the mind and excessively satiates its linguistic role, temporarily shielding it from leaky descriptions that might backdoor the mind if meaningful words were to be heard.

NDVOs should also not be too long in total duration because of the risk of fragmenting the long sound into unwanted separate strings, and so, preventing misinterpretating a word-unit from

[15] The rules on how phonemes can be combined within a language.

[16] It is worth noting that between Indo- and non-Indo-European languages the difference and similarities in suggested meaning associated with a phoneme are equally common.

a phrase. Long durational NDVOs are characterized by a high-syllable count, but can also be mono-syllabic. A syllable is described by a *pulmonic turn* or a moment wherein the laryngeal air is distinctly articulated and considered sufficiently different from the prior sound profile.

Take Stein's "Patriarchal Poetry" in section 2.2 and observe what returning does to a phrase of denotational unit. For example, how narrativity is impacted, or what happens to linearity in reading. Each further nesting amplifies the urge to resist plot or ones dependence to it. The word in the phrase loses its syntagmatic role—what does it do to other syntagma—and instigates a sonic deferral towards something intrinsic to the word. Parsing the text orally further down the syntax tree creates a presence that grows in form from a residual after rubbing against the referent. If we listen to a word, we look through it to its referent, but an insistent piece of text can make the interaction with the word two-staged: first, a return to the word demarcates and localizes it, and so we address the word and not its referent: then a second falls back on the intrinsic of the word, the heard.

Presence-by-Differential $\nabla^2 P$

Then, what does it mean to insist on smaller structures of voice both denotational and non-denotational. Iterating on clusters of NDVOs or phrases emphatically results in what I call a *presence-by-differential*. This presence P is marked differently compared to phrases built up from DVOs. I denote this presence(P)-by-differential by using the Laplace operator ∇^2, so we get $\nabla^2 P$. The Laplacian operator measures the difference between a value at a point and the average of its immediate neighborhood, or in other words, it detects how much something stands out from its local context rather than what it simply is. What I intuit is happening, is that insistence on the non-denotational creates narrative or linearity where emphasizing such structures enables them to acquire pseudo-linguistic functions. This implies that one first hears the sound and then continues to hear it more and more as an abstract language, in which forms sprout after notional, paralinguistic-morphologic differences have been exposed, iteration after iteration, a poetic experience Marjorie Perloff recounts as similar to Duchamp's concept of *infrathin*, in that the meanings emerge from the minute differences one finds in a series of casts of a mold, which she coins *micropoetics*. (Perloff 2021)

Kieran Daly, Quasi-Periodicity and $\nabla^2 P$

Guitarist and current collaborator of mine, Kieran Daly has an oeuvre that mostly deals with iterations of monophonic pitch streams that are somewhat morphologically preserved in terms of sound-shape. Instead of the 12-tone row, he aims to create a melodic framework solely by

pitch glides, where pulse salience does not have to be as dependent on dynamic articulations. His studies in non-linguistic yet highly-syntactically rich idioms and their modulations are comparable to experiments of mine in linguistic idiomatic play and especially in creating this $\nabla^2 P$ through iteration.

Taken from a correspondence, Daly describes the aim as “a structural coupling of pitch and time behavior, where the time structure is similarly gradational as the pitch structure,” where the microtonal glissando carries a quasi-periodic temporal structure also found in low-frequency jitter in speech prosody. I also work from this premise, though at the level of voiced phonation. By iterating on a NDVO a vocal presence can be created through the same local curvature, the same standing-out from a proximate neighborhood. Refering to composer James Tenney’s account of ergodic form in the article "Form in Twentieth-Century Music" (1969-70) (Tenney 2015) Daly mentions that ergodic iteration yields a condition in which “no perceptual salience is possible to obtain,” a risk he addresses through sound streams that are “heterogeneous in their adjacency with respect to one or several statistical parameters... but morphologically self-similar at higher formal levels.” The insistent NDVO also asks for same calibration: parametric deviation across iterations, or the halo collapses.

Strategic Insistence on Non-Denotational Structures

Noting that due to the lack of a denotational anchor, an ambiguity can arise as to what the constituents exactly are of the phrase (cf. word in the denotational) due to being unacquainted with what is being heard. At the very worst, the phrase is perceived as multiple phrases, or as one vocal blurb, or moving in between the previous where it is unclear about its compositeness. The proposed affordance that Stein’s concept has is that these misreadings are automatically avoided when working insistently with the phrase over and over and over again. It may now sound that whatever one chooses to repeat or insist upon can acquire this presence, which is definitely not the case. Strategy informed by context is required to condition voice to enter this stage of abstract linguistic synthesis. Without semantic grounding any insistent modulation could potentially sound as a mere repetition without any creation of differential presence. This is a pitfall I have to take into account when dealing with linguistic nonsense voicings. and especially what they become when placing them in a timeline, and how grouping principles affect them.

"TLALAATALA"

Take, for instance, the project "El libro de un libro" where Spanish poet José Luis Castillejo—member of the Milanese Zaj collective and student of Gertrude Stein—experimented on the

typographical and vocal morphologies of the letters *i* and *j* and in another project combining *t*, *l*, and *a* as heard in the phonetic reading "TLALAATALA" by Fernando Millán in Madrid (Alga Marghen) "TLALAATALA" is an exhaustion of the articulation of the NDVOs T, L, A that explores a large amount of timbral permutations of plosives, glides and vowels see excerpt 4. There is no punctuation and no spaces with sixhundred-sixty letters per page. Notably, the time intervals between iterations on the {T,L,A}-set are telling of morphological emphases on certain shapes to make the non-denotational present. Eventually, a semi-abstract and semi-linguistic sound language is created that does not want to be decoded yet acquires form through sustained, almost verbose, audible re-engagements. In a manner comparable to Inuit practices, emphasizing putting words into words, Millán enacts the letters themselves, bringing *an idea* into being through vocalization; through the mouth, he carves form out of the otherwise unbounded epistolary field of t, a, and l. Insisting on the uttered phonetic cluster suspends a presence without letting it permanently stick to the inner walls of the brain. Once this emphasis dissipates, or voice is absent, the resulting language undergoes a peculiar transformation: it shifts from the macro-scale to the micro-scale—back into philological time. During an insistent block the reader and listener is carried away from this philological time to begin with, meaning, away from the world of the book, its grammar, style, typographical body, and the hermeneutic field that opens whenever one encounters a page and asks what kind of object this is. Philological time is a micro-level temporality that is bounded, and where interpretation moves through letters, morphemes, syntactic choices, the particular decisions of a writer working in a tradition. It is the time of the book as book, and the language of the book belongs to it entirely. The voice, during an insistent block, however, operates differently. Voice makes presence, suspends the reader inside a sonorous Now, and in doing so overrides the book's own temporal logic. An insistence is a mechanism: the sustained, almost verbose re-engagement with the {T,L,A}-set accumulates force until the philological is deferred. Once that insistence ceases, the macro-scale moment collapses, and both reader and audience member return to the page—to what the text is, to the six-hundred-sixty letters, to the micro-level world the book had been holding in place all along underneath the voice.

During this insistence, a fragmentary suspension and acquisition of some presence gives rise to a temporary phonotactics and rules of the voice. Only combinations of the {T,L,A}-set are allowed and demarcate a certain zone of possible expressions, of possible abstract languages. Faced with the unfamiliar, one needs different measures for meaning in the utterances, that are based in the differences between the previous acoustic and articulatory movements and the following. A subtractive measure which becomes progressively more distinct as the letter insistence goes

on. In "TLALAATALA," perceptually enclosed phrases (e.g., mono- and multi-syllabic, word blocks)—with a clear start and end—are the insistunits, followed up by caesuras (e.g., moments of breath) to punctuate streams of sounds into vocal phrases.

To return to strategy and context, Castillejo inventively uses the concept of a book as a determinant in steering the performance of the page. He implicitly asks the text interpreter to traverse from left to right, incipit to terminus (Figure 2.2). Already faced with this constraint, it pre-conditions the reader to read it as a book. Expecting a possible narrative already frames oneself in a mode to receive sense. But this is never truly received and stays as a residue of a promise. This residue is verbalized into the NDVOs. Millàn seems to create phrases as if he is reading a text of sense. Millan's voice moves through the phonetic material with the cadence and phrasing of someone narrating a meaningful passage. A voice that is narrating a text plays on the level of the suprasegmental: intonation, stress, prosodic contours that in ordinary speech organize words into intelligible units. These features are *paratextual* in Genette's sense, framing and shaping what surrounds the linguistic content itself. (Genette 1997) And in this performance, this voiced residue of a book-reading—the habituated prosody of someone reading aloud—is definitely suprasegmental and is superposed on the wild barrage of *t's*, *l's* and *a's*. The book-form, carried in the voice as cadence, presses the gibberish into shapes that the ear receives as seemingly syntagmatically coherent. Units seemingly belonging to a natural language, even without denotational content. In my work for synthesized voice *PP: Phonesthemic Palimpsest* (2025), a sentence begins as one large string of phonetically encoded characters and acquires sentence-form through the application of phonotactic rules such as words of justifiable length, correct placement of definite articles, phonotactic sequencing. In what can be called a paratextual exercise, I aimed to produce coherence in an imaginary language just by some imagined grouping principle alone, with no voice to carry it.

Furthermore, Millàn is Spanish and enunciates the text with a Spanish tongue-palate/palette play. Therefore, the recitation of this text is already influenced by the articulatory motorics of the native language of the sound poet. In some regard, the phonotactics of this nonsensical language created by the t's, l's and a's are already twisted. The way 't' sounds after an 'l' is different when read by native Dutch speakers. Further inspection shows that this specific consonant cluster /tl/ is rarely allowed in Spanish phonotactics as an onset. However, /tl/ is perfectly legitimate and easily pronouncable by the Spanish speaker, and is commonly located between or at the end of a

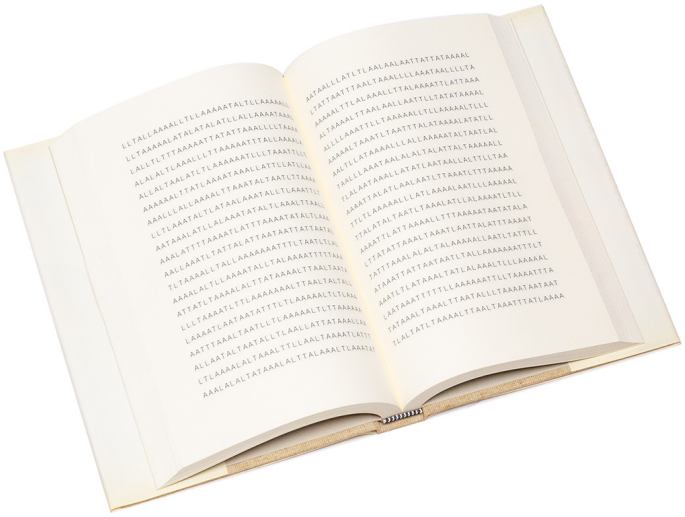


Figure 2.2: Two pages out of "TLALAATALA"

word.[17][18] Millàn tries his best to not fall into a habitual mode of phonation, such as unconsciously favoring /tla/ because these literally roll off the tongue in comparison with, for example, /lta/, but eventually does let his tongue roll and let 't' follow 'l' more often. Articulatory phonetics explains this simply because /tla/ matches the natural timing and coordination of articulatory gestures better than /lta/. The tongue tip stays engaged and simply changes shape in /tla/, but in /lta/ the tip must disengage, reposition, then re-engage, which is slower and more complex. On the inscription level, the author, a Spaniard, might have an orthotactic bias when writing this text.[19]) Nonetheless, Millàn is the voicer of the text and the text's formation of phonotactic order is mainly generated by him, and assimilated synchronously by the auditor as a sound structure reminiscent of a personal language. The text itself is generative, but it is the orator who eventually cuts out form from the otherwise almost unending stream of characters.

The goal is to achieve a reading of "TLALAATALA" where the enclosing of the word and phrase is perceptually reliant on the auditor. That the formation of an NDVO is up to the listener's perceptual reception. Here, a synthetic reader would avoid factors contributing to orator-bound linguistic segmentations. A digital text-interpreter can be built to parse and synthesize the text

[17] If a Spanish word starts with /tl/-, there is a very high chance it is a Nahuatl loanword or a proper noun.
[18] Arguments of the sort: *possible sound x changed to y in language A because that language did not have (did not permit) an x* are circular and reduce to the tautology; lack of x because of lack of x
[19] A preliminary analysis on a small text corpus from "TLALAATALA" yielded a significantly higher frequency of /tl/'s than /lt/'s.

without much hermeneutic noises coming from the primate.[20] This interference can come from the innate faultiness of the vocal mechanics and the human's cognitively biased mediations. The necessity of breathing limits the length of voice phrasing. The synthetic verbalizes the text uninterrupted and unsegmented. This parse-like continuous voicing asks for a private and phonetic closeness to the objectively distant cacophony and simultaneously arranging it into a logic endemic to the listener's ear.

"Automaterga"

Another example of what form an NDVO can take on and specifically how presence is given to iterative engagements with the nonsensical can be seen in the computer generated poems "Automaterga" by Dutch poet and flutist Greta Monach. Over the past years, I have read, performed and generated some of these myself. As with Castillejo, each "Automatergon" makes use of sets of NDVOs, separated into three categories; onsets, nuclei and codas. In contrast to Castillejo, what character likely precedes or follows another in an Automatergon is defined beforehand, and the final step of writing the poem or text is done algorithmically opposed to cerebrally, instruction after instruction. NDVOs are presented to the interpreter as monosyllabic strings contained within cell of a rectangular grid. There are clusters of cells (depending on a density parameter) exploring timbral variations on one syllabic constituent (i.e., onset-nucleus-coda). It is up to the interpreter to improvise a spatial choreography within the grid where the left-to-right reading paradigm[21] is just one possibility of traversal among many. Two automaterga *72-24B var. 1* and *72-25A var. 1*—performed by myself in sequence (5)—investigate the orthogonality of the grid by vocal parametrization of the vertical Y and perpendicular X-axis, she mentions that "...relations between words become active in 2 dimension, which means in many directions" and emphasizes that interpretations should vary "with the wanderings of the eye over the page." (Lansdown 1973, 7) Compared to "TLALAATALA," the insistunit happens on a morphemic level and not on the phrase. This suprasegmental residue found in Catillejo is not heard, partly due to a lack of variance in the word lengths that makes the strings stay perceptually distinct. It becomes a study of iterative efforts on a similar voice sound shape, rather than depending on the rhapsodist to create order. This makes the reading viable for direct insistence on any voice shape of a shorter duration and low information density. Rip out one page of Catillejo's book, and the monospaced characters immediately become something else: the previously mentioned incipit or

[20] This is not to claim that the computer does not interpret, it most definitely does, rather why it matters here is because a computer program is always instructed to do exactly what is written down.

[21] What comes closest to this type of reading is the 19th-century letter writing technique where an author fills a page, rotates the paper 90 degrees, and writes a second layer of text directly over the first; also called *cross-writing*.

first word becomes more a visual boundary than anything else, and makes the text from the previous and next page(s).[22] Who stops the sound poet to treat one page in "TLALAATALA" as an "Automatergon"?

Voice Sheafs and Non-Denotational Structures

These two verbal scores, seem to let the reader tread water only when continuous vocal engagement is made. Both the voiced and written are barely keeping their head above water, only propelled upward by the paddling of the proverbial feet (e.g., the tongue) and sparse thrusts of buoyancy coming from accidental, linguistic, salient sounds: where signifying audio figments like ideophones or even words can float to the surface. The sounds become entangled in an abstract form, only for a moment, to then drown out in the sea of associations voice and text usually bring upon their hearer and reader. The texts share an affinity in visual and phonetic form, yet some jargon or framework at our disposal is needed to describe NDVOs that—in the broadest sense—share data in a (pseudo-)linguistic-acoustic space. Some abstract structure that captures a vocal unit (word, cluster, or phrase) that has a stable morphological identity while remaining permeable at its edges: transformed, projected, or mapped onto other such units while perceptually preserving something essential is required. An option for a model can rely on both psychological models of categorization, which would suggest describing vocal events experientially, done so in electroacoustic contexts. (Bergsland 2010; Wishart and Emmerson 1996; Calderón-Garrido and Gustems 2015; Smalley 1997) Aligned with the ethos of this chapter I chose to rely on a model of abstraction that takes a stance against idealizing a necessity of the *noeme* ("that-has-been") in order to achieve *expressivity in voice*. Then, the construct that I deem most fitting to describe units of non-linguistic vocal expression is the mathematical object called a *sheaf*[23]. Before conceptually marrying this mathematical object to voice, several terms need to be unpacked, but most importantly, a clarification has to be made about its scope. It is an abstract model and naturally does not explain voice perception, but its function is as analytic as it is creative from the position of the sound poet. Especially proving useful that this tool's *language of description* originates from axioms, and goes against the naturalness of Natural Language; a natural essentialism I would like to avoid in all things voice. We are using the language of the Topologist to name relations that already exist between NDVOs and to not give a causal account of why those relations might arise, id est, trying to *explain* perception.

[22] If not treated as a book, then these paratextual/suprasegmental effects are not impeding the reading if the intention of the reading is a formal study on the sound of voice units.

[23] Most of my research in sheaf-theory comes from an introductory book called "Sheaf Theory through Examples." (Rosiak 2022)

A sheaf is a way to organize *local data* over a topological space in a way that is consistent across overlaps. More precisely, there is a topological space (i.e., the stream of voiced sound) that over each open set (each perceptually coherent and distinct region) there is some *data*, some structure, information, object that lives over that region. The space, plus the data over every open set, plus rules for how data on big regions relates to data on smaller regions inside them is packaged in this sheaf.

The topological space is a set of points together with a specification of which subsets are called open. It is a minimal way of formalizing a neighborhood and closeness without requiring a numeric distance. The rules are spare: the empty set and the whole space are open; any union of open sets is open; any finite intersection of open sets is open. An open set has the property that every point inside it has a small region around it that is also entirely inside it—no point sits exactly at an edge with nothing around it. A closed set, by contrast, contains its boundary. An open set holds all points satisfying $x < 1$, a closed set holds all points satisfying $x \leq 1$. A set can also simultaneously open and closed or *clopen*. A clopen set can be useful, for instance, to describe a voiced phrase that includes its silences (closed, bounded) versus one that bleeds into surrounding material (open, without hard edges).

I will treat the perceptual-temporal continuum of a vocalization as a topological space. Where a point is a voiced, acoustic event, and more specifically, in this chapter, an NDVO. Then, a phrase in, for example, Millan's *TLALAATALA* is an open set, defined as perceptually coherent regions where the voicer and listener group acoustic glue events together. Their boundaries are not sharply demarcated; one can be more or less inside a phrase depending on how legible the imaginary sentence-divisions articulated by the reciter are. Data is to be found over each open set of this space, meaning each phrase-region, word-region, and syllable-region. The phonological-morphological identities of that vocal unit are made up of spectral profile, articulatory shape, intonation contour, and its felt vocal presence.[24] A *sheaf* is the total organization of these streams of information: the space, plus all identities assigned to every phrase, plus the rules governing how identities on large regions relate to identities on smaller regions nested inside them. These rules work through what are called *restriction maps*. If you have a *section*—a specific phonological identity assigned to a phrase-region—you can restrict it to a smaller open set inside that region, like a single morpheme or syllable. This restriction is the formal act of asking: what does the phrase-level identity say about this sub-region? A *local section* is then some identity assigned to one particular open set, but not the identity over the whole space.

[24] These categories are called presheaves.

Finally, a *global section* is a consistent assignment of identity across the entire space or topoi in one go. How the sheaf comes in handy is that it assigns local data to every open set of a space. How local connects to global is described by an important axiom called the *gluing axiom* which roughly says that compatible local sections always assemble into a unique global section. The logic is similar to a jigsaw puzzle. Once the edges match, there is exactly one way to fit the pieces together. If you have local sections or *functions* s_A and s_B over a collection of open sets A and B where those local sections agree wherever their regions intersect ($A \cap B$), that is, the restriction of s_A to the overlap equals the restriction of s_B to the overlap; then there exists a unique global section over the whole larger region that restricts back to each local section. Compatible local pieces can be glued into one coherent whole, with exactly one way to do it.

What makes Castillejo's text and Millán's recitation relevant the discussion is that this gluing axiom is only weakly satisfied. Sometimes, adjacent NDVOs share an acoustic overlap with their release trail bleeding into the attack of the next voicing—but their phonological identities do not fully agree on that overlap: the shared region sounds slightly different depending on whether you approach it from the left unit or the right. The local sections are compatible enough to suggest a global section, phrase, or presence, but these never quite glue into one. A unique global reading never really crystallizes. Instead, a family of candidate globals hovers, and the listener is eventually left doing the final gluing layer themselves. The presence produced by insistence is asymptotic in affect and forms somewhat endlessly into a global section. Greta Monach's "Automatergon," on the other hand, handles a different sheaf logic. The gluing axiom is structurally inapplicable, not even weakly satisfied. To understand why, it is necessary to look at how the poem's generative logic programmed to then connect it to a sheaf its structure.

Every word in an Automatergon is built by a *wordGenerator* function as a strict concatenation of three phonological slots[25]:

```
word = onset + vowel (nucleus) + coda
```

The output uses Dutch phonotactics as seen in figure fig. 2.3. Every word in the upper-left cluster ends in *-ag*: gnag, smag, schrag, pag, spag, tjag, blag, snag, frag, flag. The top and bottom-middle cluster ends in *-og*: blog, og, nog, sjog, slog, spog, sprog, schroog. The lower clusters shift to *-ig* (blig, slig, snig, schrig, sprig, wig) and *-ug* (prug, blug, flug, schrug, gnug). This is because the coda is chosen once per poem and held fixed for the entire grid. It is the one element that is globally constant.

[25] An implementation of the Automatergon poem can be found here `https://neuskeeloor.app/automatergonJpKrNe_2dslider.html`.

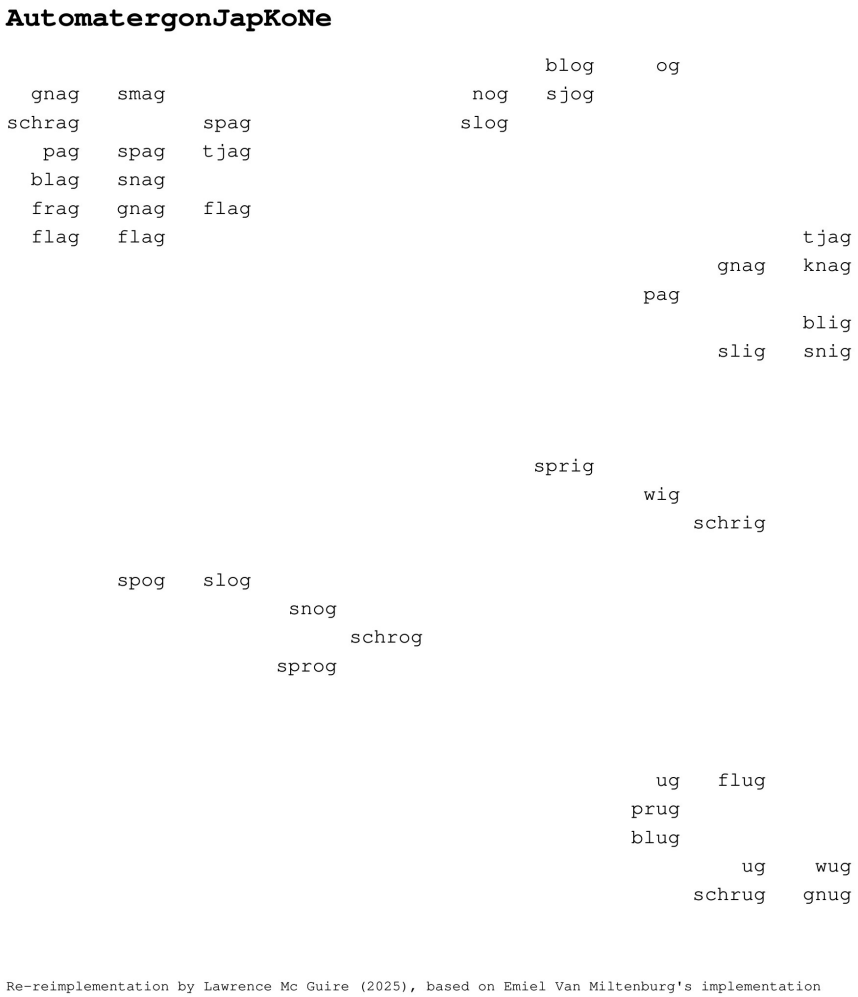


Figure 2.3: Automatergon with fixed coda, varying inter-cluster nuclei, and intra-cluster variation onsets

The onsets and nuclei in figure 2.3 are not fixed globally. An *optionMap* function constructs a set of categories, either organized by vowel (all words in a chain share the same vowel, onset varies) or by onset (all words share the same onset, vowel varies). When a chain is placed in the grid, a single category key is selected at random and held for the duration of that chain. So within the upper-left cluster, the shared vowel is *-a-* and the onset varies: gn-, sm-, schr-, p-, sp-, tj-, bl-, sn-, fr-, fl-. The timbral identity of this cluster comes from the locked vowel and coda, with the onset providing variation within a phonological neighborhood.

Spatially, chains are grown by the *getAvailableNeighbors* function, which checks all eight surrounding cells (orthogonal and diagonal adjacency). A chain starts at a random empty cell, places a word, then steps to a random empty neighbor, and continues until the chain length is exhausted. This produces irregular but spatially contiguous blobs, or groups of cells that cluster together with empty space between them. Each cluster is, in precise topological terms, a *clopen* subset of the set of filled cells that is importantly, structural and not incidental.

Consider the filled cells of the grid as a topos where the open sets are defined by adjacency: a set is open if every cell in it has at least one neighbor that is also in the set. A chain-cluster is closed because it is a bounded, spatially isolated island. The empty cells surrounding it form a definite boundary, and the cluster contains that boundary in the sense that no cell bleeds into an adjacent cluster. It is simultaneously open in the subspace of filled cells, because every interior cell of the cluster is surrounded by other filled cells of the same cluster. The empty space between clusters is different to the silence in Castillejo's sense where the difference is that the space acts as a physical topological separator, the gap that makes each cluster clopen: disconnected from every other cluster, internally connected within itself.

This is different from Castillejo's open sets, which bled into one another through some acoustic overlap. A separateness between clusters is built into the algorithm with some rules jointly stating that a chain cannot enter an already-filled cell, so that clusters never merge. A cluster is a complete, bounded phonological world to itself. The onset-nucleus-coda structure maps directly onto three nested levels of local section, each with different degrees of fixity. The coda holds everything together in this example and is the global section. And again, in Tristan Garcian's concept of the *thing* and the *world* terms, it is the closest the poem gets to a *world*-condition, the one element that is in every thing without being a thing itself. (Garcia, Ohm and Cogburn 2014) Every NDVO in the poem, regardless of cluster, shares the same closing consonant. The gluing axiom is satisfied here: if you restrict any cluster's identity down to its coda, you get the same answer from every cluster. The global section over the whole grid exists and is unique, but

it is thin and carries only one bit of phonological information.

The nucleus (vowel) is a local section over each cluster. Within a cluster organized by vowel, the nucleus is fixed and functions as the cluster's identity—the shared feature that makes the blob cohere as a recognizable phonological neighborhood. Restricting this section to any individual cell within the cluster gives the same vowel. But restricting the section of one cluster to the empty boundary cells and asking what the neighboring cluster's section says about those same positions yields nothing, or rather yields incompatible answers because the neighboring cluster has its own independently chosen vowel. The gluing axiom fails at the inter-cluster level: two adjacent clusters cannot be glued into a larger coherent section because their local sections disagree on the shared boundary (the empty cells between them). Strictly speaking, the clusters are clopen, separated, and the gluing has no overlap region on which to check consistency.

The onset is the most local as it varies freely within each cluster. It pops up as the noise within the local section, the timbral permutation that prevents the cluster from becoming a pure repetition. It corresponds to what was earlier described as the mechanism that keeps NDVOs perceptually distinct while remaining morphologically related: the onset is what makes gnag different from smag different from schrag, while the shared *-ag* keeps them audibly in the same family. In terms of differential presence the sheaf logic in the "Automaterga" diverges from "TLALAATALA". In Castillejo, the open sets overlap. The gluing axiom is weakly satisfied so that adjacent local sections are almost compatible, and the failure of full gluing produces the briefly mentioned halo-like presences, global sections perpetually forming. The listener does the mental gluing work in the gap between local sections. The local sections in the "Automaterga" do not even attempt to glue. The clusters are clopen: they have no overlap region on which consistency could be checked. The gluing axiom is simply inapplicable, because the precondition for gluing (shared overlap) is algorithmically and structurally prevented. This tells perception-wise that the interpreter is not invited to do gluing work across clusters. Instead, each cluster presents a complete, self-contained phonological identity that closes on itself. The presence produced is not the halo of an arriving global section but something more like a series of isolated islands related to the others only through the thin global section of the shared coda. The coda thus functions as a morphological skeleton. A minimal global coherence that is audible in every word but never suffices to make the whole cohere into something more than its parts.

Insistence: Thickening The Vocal Stalk in Denotational Structures

The claim is that, within this sheaf framework, insistence does create a situation where *weak gluing* takes place between NDVOs. Sheaf theory has an additional structure called a *stalk* that helps describe how insistence on DVOs plays out concerning the time model of density and saturation. In *sheaffian* terms, the stalk F_x at a point x is the colimit of all sections over shrinking open neighborhoods of x. Translated: it is the data that remains once zoomed in infinitely close to a single moment, called the "germ" of the section at that point. Nothing of the global sections' behavior matters for the stalk; only the infinitely local does. Stein's insistence, read through this, is precisely a technique for *making the stalk rich.* The stalk is already there, untouched and not activated. Each return to the word or phrase is a new local section over a slightly different temporal open set, as in it frames the *same* vocal event differently. The more one insists, the more one carves into the morphological core (the word's intrinsic shape and phonetic identity). These sections agree on this level, which is why they feel like the same word, and yet these sections still don't completely agree with one another due to being different at the edges. Different as their emphasis slightly shifts, the syntactic surround varies, and the listener's anticipatory state changes. The iterations are local sections over overlapping but non-identical neighborhoods of the word's "moment of presence." Insistence functions as a way to build a thick stalk with a germ so granularly specified by repeated local approximations that the spoken word takes on an infinite local density. It is the diametrically opposed act of "hollowing out" the word, or "fully carved out essence," or through Pynchon's idea of temporal bandwidth and its compression Δt into a saturated point. Insistence cultivates attention inwards, away from the global section, and away from the coherent reading of the whole and narrative.

Considering these attempts at understanding the dynamics between both NDVOs themselves and DVOs and its structuring over time, space and phonology gives us to some extent what we need. But there is still something that has to be dealt with if one truly wants to approach voice in a non-denotational way. Both Monach and Castillejo work relies on the coded nature of language to then push towards experiencing a sonic form rather than one to do with speech, and arguably are not working on a deeper issue regarding codes in voice. Human vocal anatomy is preconditioned to work within a certain kind of way, in certain ways of movement, in a morphologically constrained way. Therefore, phonetics is the next issue that will be tackled. Insistence will once again prove necessary here as a sound-attention conditioning device.

2.5 Phonetic Dependency and Coarticulation

Speech can be seen as a system of patterning voice where sound patterns are sonically variational depending on its phonetic and phonemic environment. Phonology is the study of these patterns and especially the relational logic that governs which distinctions count and which do not. Jakobson and Halle's describe in their landmark work "Fundamentals of Language" that the minimal unit of phonological analysis is not the phoneme as a discrete sound but the distinctive feature. The binary opposition such as voiced/unvoiced or nasal/oral acquires meaning only through contrast with its counterpart. They describe the listener parsing a continuous sound-stream into "a definite number of successive units," each presenting "a definite number of paired alternatives used with a differentiating value." (Jakobson, Halle and Halle 2010, 13) A distinctive feature carries meaning only in context with others, and therefore phonology can be seen as a grammar of phonetic patterns within a language. What is allowed to be contrasted with what; what combines with each other? More concisely: which sound configurations are licit and which features of one segment bleed into or constrain the features of its neighbors? In some form, one can talk about phonetics in a similar tone. Sean Gurd describes phonetics as "...that layer of language in which are organized the physical gestures which manipulate the voice" (Melnick 2014, 12), literally pointing to voice as being submissive to the phonetic system, both neurologically and anatomically it acts as a mask. Both fields are systems dealing with cases of fossilized constructs. Defossilizing these means going against the anatomically encoded aspect of voice in all its possible modes: speech, song and non-standard settings.

According to Dutch psycholinguist Willem Levelt, known for the Levelt model of speech production, speech production has a processing pipeline: after the stage of Conceptualization(thought) and at the end of the Formulation stage(choose best lexical canditates), a phonetic plan (also called the articulatory score) is set up and provided to the Articulator, after which the Articulator turns these syllable-sized programs into neuromuscular commands making use of a mental syllabary; similar to consulting a stored lookup table of pre-compiled motor programs, or like a mental lexicon, but for gestures rather than words. For a high frequency syllable like /ðe/, a fully pre-compiled gestural score is retrieved[26], but for a novel syllable such as a nonce word there is no stored program and the Articulator has to synthesize a program from component gestures, which is why unusual phonotactic sequences often cause for vocal trip-ups. Psycholinguists Cholin, Schiller and Levelt confirmed this with monosyllabic pseudowords in another study: high-frequency syllables are produced faster and more fluently. (Cholin, Levelt and Schiller

[26] From the English-Corpora: COCA.

2006) It is safe to say that this retrieval advantage would almost certainly bias random generation of words or even sounds. From the random number generation literature the human behaves systematically non-random when asked to produce sequences "at random" as they fall back on whatever is least cognitively loading.(Myachykov et al. 2020) Since high-frequency syllables have a processing advantage at the level of phonetic encoding, the path of least resistance during random generation would favor them. This tells us that even "improvised" phonetic gestures such as NDVOs and nonce words contain motoric traces from fixed, encoded targets—taking on the form of articulatory interpolations. More generally, the production process of voice—even when not dealing with overt speech—is equally fossilized, determined. Yet, knowing the proverbial launch sequence for any word does not fully explain how a signal is perceived as being vocal. This issue is called the "lack of invariance" in speech perception. The same phoneme /d/ takes on a different spectral shape in *dip* than in *dog*, it also varies across speakers, rates and coarticulatory environment, and yet is perceived as categorically identical. First foregrounded by Liberman in 1967, in that the listener is actually decoding an articulatory intention, meaning that the acoustic waveform has no one-to-one mapping to the phonological unit. The signal underdetermines the percept. Liberman then came with a theory to partially resolve this problem, named the "Motor Theory" of speech perception. The percept *is* the intended gesture and not the sound, meaning the listener recovers the speaker's articulatory intention by means of a shared motoric substrate: the same system that produces speech also decodes it. The stored articulatory score from the mental syllabary functions as the inverse map from signal back to articulation.[27] Current research in the activation of the physical voice mechanisms by non-speech sounds such as hand claps or environment sounds show no relational effect. (Rogers et al. 2014; Michaelis et al. 2021) The scientific literature shows little interest in those sounds embedded with a more voice-like temporal and spectral modulation[28], but still not clearly perceived as speech, and how these can drive locomotion of muscles (e.g., noise-vocoded speech or chimeric stimuli). This is where the sound poet comes in to turn this gap in "Academic Knowledge" into an opportunity to displace the laboratoric voice—always meddling in scientific doxa—to a another environment: the one of music. The phonetical system is a mask, or something that constrains voice to a bounded field of expression[29], then the motoric substrate that enforces this constraint is also, necessarily, the substrate through which any voice, artificially or naturally produced, will be decoded, quoting

[27] Evidence for excitation of vocal muscle activations are shown in measurements of movements in the vocaltract during passive listening experiments to specific speech sounds. For example: lingual consonants in utterances (Fadiga et al. 2002); seeing and hearing speech sounds enhances tongue muscle activity. (Watkins, Strafella and Paus 2004)

[28] A pitch contour, rhythm of speech, timbre, movement within the timbral space or formant field.

[29] Reminiscent of Inuit thought, this would be considered as the form bearer, the carving tool giving form to unbounded of reality.

Dada poet Tristan Tzara that “thought is made in the mouth.”

I want to state that listening to voice is as auditory as motoric and that it is doubly conditioned. Eying Fowler’s theory of Direct Realism, the gesture is not inferred but directly perceived; yet the auditory cortex remains the medium through which this gestural information first arrives—and is processed—hence vocal audition is simultaneously a pickup of distal articulatory structure and a processing of proximal acoustic shape.

Again, I consider voice production and perception to be rooted in the same motoric substrate. The vocal tract can be considered a bounded morphological space, and what it can produce is a set of constrained zones of possible shapes. Saying that our substrate is not infinite. Long have experimental vocalists dealt with the anatomical and biomechanic inertia by relying on acoustic and extended bodily tricks to suspend voice to produce non-standardized voices. Often finding themselves to push against the physical limits of what is possible, and after enough practice eventually moved these boundaries. But still, the possibilities are delimited, even praxes of *embodied becoming* bump into the pharyngeal walls.[30] The motor programs that build up the mental syllabary are samplings of that space, which together form a sort of field of vectors that Natural Language historically chose to pragmatically stabilize. Moreover, one articulatory movement implies a synchronously coupled movement of another. This coupled movement is *phonetically dependent*. This fixity or interdependence of the substrate on anatomic and neurophysiologic level can become newly productive by clarifying some things. A statement that continuously roams my mind once voice pops up is the notion “voice through other means.” First of all, asking “what is voice?” and consequently “what does it do, and to what does it do it to?” might be a good beginning. Christian Clozier titles a polemical in Henri Chopin’s *Poésie Sonore Internationale* with "VOIX ET ELECTROACOUSTIQUE" which has stayed with me more as voiceprint than ortographic, hearing “voix *est* electroacoustic,” translated “voice is electroacoustic” instead of “voice and electroacoustics.” Following the advent of phonographic revolution of magnetic tape as, Clozier says, voice has become[31]:

> ... not only speech or song; it is everything that comes out of the mouth, everything that is too weak to come out and that one seizes with contact microphones; it is the natural voice in its acoustic space, the amplified voice, the telephonic voice, the

[30] Embodied becomings as in Judith Butler’s account of the body as materialising through repeated performative acts rather than expressing a prior essence, or crip theory’s revaluation of non-normative corporeal futures as sites of agency rather than deficit the vocal apparatus are not thereby rendering the voice as topologically open. Extended vocal practices are genuine instances of crip *self-fashioning* at the laryngeal level, and are real expansions of what the *performed body* can do (see Butler 1999). But, however affirmative this morphological freedom, it still operates within a bounded space, only remapping an interior.

[31] My own translation from French to English.

> manipulated voice, the voice carried on the airwaves, the voice of other peoples and other cultures. (Chopin 1979, 277)

Today, voice is primarily experienced in a transduced state, arriving to us in mass-mediatic forms through technologies of the written and the sonorous. Clozier writes, "the myth of voice is dead. It is no longer the voice that makes the human; it is the human that makes their voice." Before phonographic technologies of reproduction and dissemination, voice was understood as an essentially human phenomenon—chained by the body and exhausted in the moment of utterance. But after their arrival, this self-evidence collapsed. Voice(s) can now literally leave the body and—through phonographic inscription—inhabit cultural devices that transform, distort, enhance, spread, resound, revoice, clone. Once voice is displaced from body to medium, the vocal identity sustains to some degree as a hauntological product. The perceptual walls of the bounded morphological space described above still stay strut when hearing the digitally synthesized, or electroacoustically transduced voices. These *voices through other means* produce vocal percepts that are decoded by the listener through the same motoric substrate that Motor Theory identifies in the reception of natural speech. Again, the origin of the signal is irrelevant to the decoding apparatus, giving priority to detecting whether the signal has some gestural information traces.

I put natural and synthetic voice on the same plane of vocalness. And only differ in how strongly they activate the listener's motoric decoding. This decodability becomes the criterion for measuring sonic vocality. To some respect, I deem the synthetic voice more analytically transparent than the natural, because it permits the independent manipulation of parameters that in natural voice are phonetically entangled, coarticulated. Synthetic voice can tell us as much as natural voice about what *the vocal* is.

2.6 `NKOAPP`: Simulating Motoric Independence in Voice Production

To work in and against this bounded space seriously, whether in analysis or dealing with its composition, it requires modeling voice at the articulatory level at which these constraints actually manifest themselves. Bottom-up instead of top-down, or in other words, if I would base the model on the acoustic waveform and all of its derived forms of signal representation then the only decisions possible are these downstream which is not what is needed.[32] Also, modeling

[32] The Spectralists based their compositional aspirations on using downstream models (Static Spectral Modelling and Dynamic Spectral Modelling) to intervene in the identities of sounds, but being downstream was where they liked to be, where they wanted to use known voice to extend onwards through methods of resynthesis and filtering.

at the phonological level essentially deals with representing voice as a sequence of categories and their combinatorial rules. Both linguistic annotation schemes and text-to-speech systems have this fixity invisibly baked in and abstract away this coarticulatory entanglement. What is needed is a model that represents the sounding source and adheres to Natural law of physics. And, when mentioning gesture, I am referring to time series such as articulator trajectories, changes in supralaryngeal constriction positions and angles, and contours of acoustic parameters (i.e., fundamental frequency, loudness, etc.). The program developed during this period of the project makes use of a highly expressive digital physical model of the voice in the VocalTractLab software. (Birkholz 2013) Essentially it is able to perform all the tasks dealing with the kinematics of the vocal tract listed above. I introduce two ways of liberating voice from its phonetic dependencies using this synthesizer.

The first approach is driven by starting from a blank slate so that the coarticulatory constraints are bypassed. The idea is to generate synthetic utterances programmatically by way of algorithmic processes or direct manual inputs to control phonetic features independently.[33] Each gestural parameter can be treated as if it were independent of its neighbors in time.

The other approach relies on harnessing this coarticulation extracted from a found sound registration. Accessing these coded movements means having an efficient way to retrieve an *articulatory blueprint* and use it as a point of departure. It must capture the following aspects in the waveform:

1. continuous interpolations between utterances
2. overlap schedules between articulators that move at different rates
3. phase relationships "..."

The articulatory blueprint can be split into eleven control parameters for the vocal folds model and nineteen for the vocal tract state in table 2.1. All of these values refer to some physical location, translation, rotation or shape, but we are concerned with their behaviour over time and in how they manifest durationally.

2.6.1 Approach A: NDVO Synthesis

Recalling back to requirements an NDVO must meet, so must I approach synthesizing them. Exercising prudence in deciding if the vocalization will contain a global oscillatory behaviour (i.e., reduplication) is paramount. Also, a split of the source NDVO into two or more NDVOs

[33] As in: place of articulation, manner of articulation and phonation types.

Table 2.1: Control parameters for glottis and vocal tract in NKOAPP

Vocal Tract	Glottis Geometric Model$_{GM}$
Horz. hyoid pos.(HX)	f0
Vert. hyoid pos.(HY)	Subglottal Pressure
Horz. jaw pos.(JX)	Lower displacement
Jaw angle (JA)	Upper displacement
Lip protrusion (LP)	Chink area
Lip distance (LD)	Phase lag
Velum shape	Relative amplitude
Velic opening	Double pulsing
Tongue body X	Pulse skewness
Tongue body Y	Flutter
Tongue tip X	Aspiration strength
Tongue tip Y	
Tongue blade X	
Tongue blade Y	
Tongue side elev. 1	
Tongue side elev. 2	
Tongue side elev. 3	

due to exceeding some conditions for morphological continuity (e.g., being too long in duration; a silence punctuation, etc.) has to be accounted for. The idea is to generate articulatory states not yet indexed or logged in our motoric substrate, and to then move between them in non-coarticulative ways. When this NDVO is created, it defines what I call *sheaf-A*. An organizational structure over its internal temporal extent, whose *global section* captures its spectral profile, articulatory shape, intonation contour, and acoustic impression. The insistent modulations consist of iterations which agree with the global section over most of the NDVO's internal extent but deviate at specific sub-regions. In short, an iteration is an *alternative local section* (*ALS*) of *sheaf-A*. Sequential *ALS*s condition the listener to pay attention to the infrathin, the accented shape, perceiving its global difference in vocal shape and cut the formlessness into the abstract sense-percept. Sound excerpt 6 shows ten iterations on a insistunit of around 3 seconds. Its *sheaf-A* or collected base shapes is shown below in Figure 2.4. Further info on how to use the Vocal Insistunit Synthesizer (VIS) is given in Appendix 3.

Each ALS is a perceptually distinct mental entity, marked by a clear beginning and end. It does not matter how long the "silence" is between each *ALS* because each instance should carry enough data that agrees with *sheaf-A* so that it becomes self-evident they belong to each other, no matter their spacing. Silence here is regarded as:

> every sound or feature of sound not agreeing with *sheaf-A*

The cases such as hearing multiple NDVOs being insisted upon at once or hearing them in combination with DVOs will be discussed shortly in my text-sound composition *Four Saints in Three Acts* (2026). For now, the *ALS* are heard in sequence, such that onset and ending sound never meet. A collection of *ALS* in sequence together form the voice *stalk*. It is the collection of

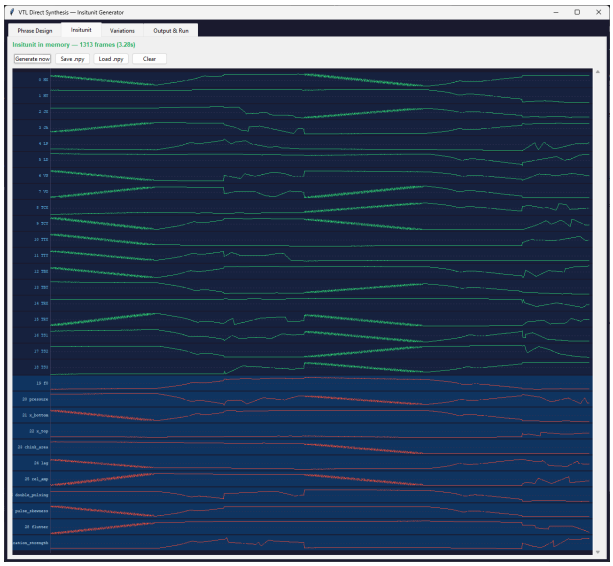


Figure 2.4: *sheaf-A* of a NDVO generated by the Vocal Insistunit Synthesizer (`VIS`)

local sections all converging on *sheaf-A* without any two of them touching, separated by silence, each carrying enough agreement with the global section to keep the stalk coherent as a single accumulating presence-by-differential $\nabla^2 P$. Figure 2.8 shows a visual abstraction of the vocal stalk over three stages. Imagine a blank paper and drawn on it a scribbled line. The scribble represents a two-dimensional projection of a *sheaf-A* onto the paper. Stage (i) makes another scribble or *ALS* agreeing with *sheaf-A* over most of its extent but deviates locally. Stage (ii) shows ten similar scribbles tracing the same general shape of *sheaf-A*, and stage (iii) shows 20 scribbles. The region that gradually saturates becomes the stalk. The accompanying sound examples are 7, 8, and 9, each playing back each *ALS* in sequence except for 7 because of being *sheaf-A* (Figure 2.5; Figure 2.6; Figure 2.7).

It is possible that the resulting *ALS* can agree with a phonologically coded gesture or at lease overlap with some correlated movements. Stage (iii) in figure 2.8 shows how—through iterative efforts—overlaps with the natural phonetic regime can happen and simultaneously define the stalk more and more clearly.

2.6.2 Approach B: DVO Resynthesis and Intervention

Knowing a priori how motor control must be operated to achieve a desired acoustic result is a control problem fundamental to articulatory-speech synthesis (Krug et al. 2023, 1), but one that has recently seen garnering new interest in fields of Neural Control.[34] After making the phonetic dependencies legible as structure and able to be parsed and rendered by `NKOAPP`, we intervene in

[34] Grouwels, Jonason and Sturm[2025]; Krug et al.[2025]; Cámara et al.[2023] &Südholt et al.[2023]; the evoc-learn project (Xu et al.[2024];Niekerk et al.[2023];Van Niekerk et al.[2022])

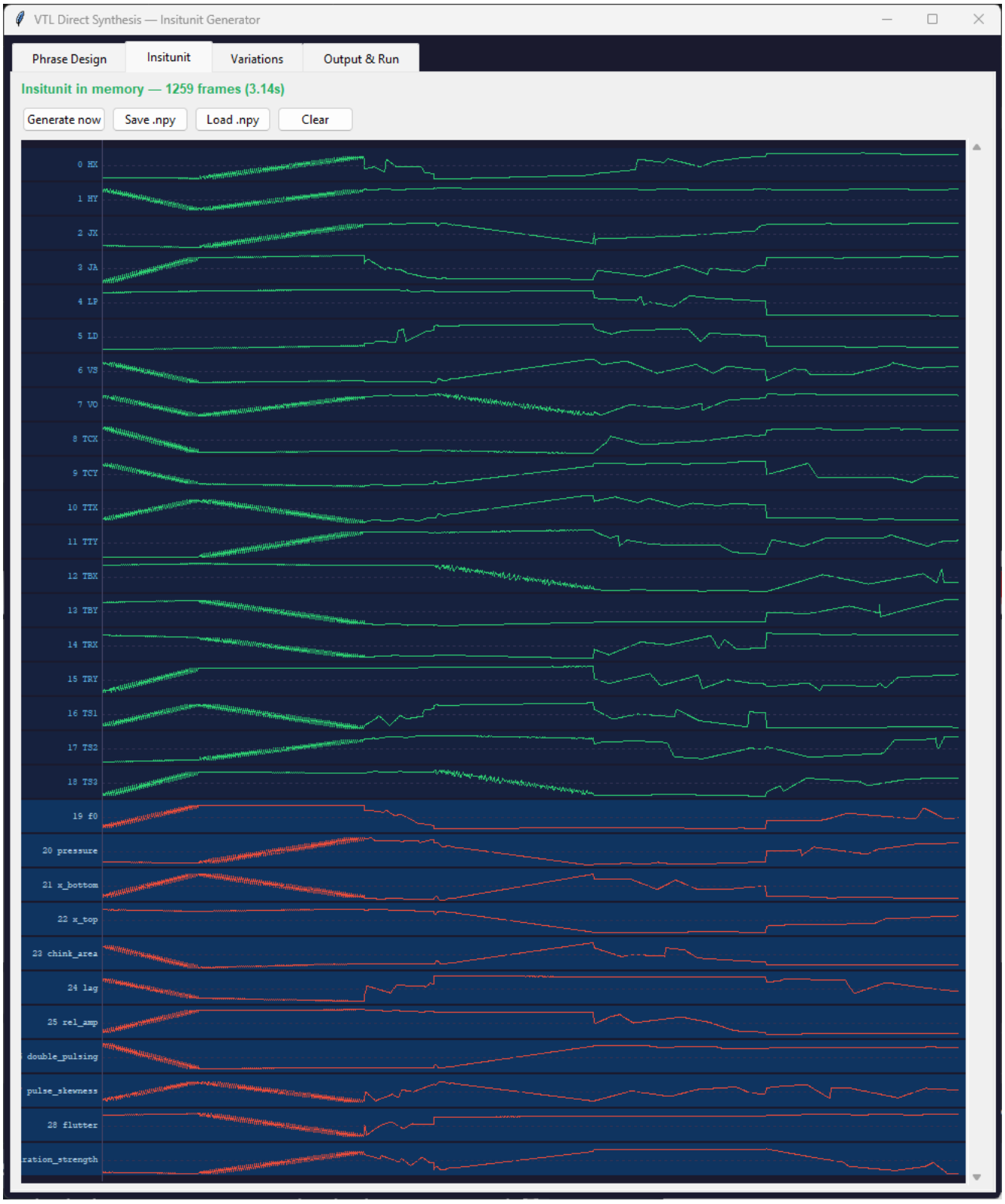


Figure 2.5: (a) *sheaf-A* or the base shapes of (i) in `VIS` see fig. 2.8.

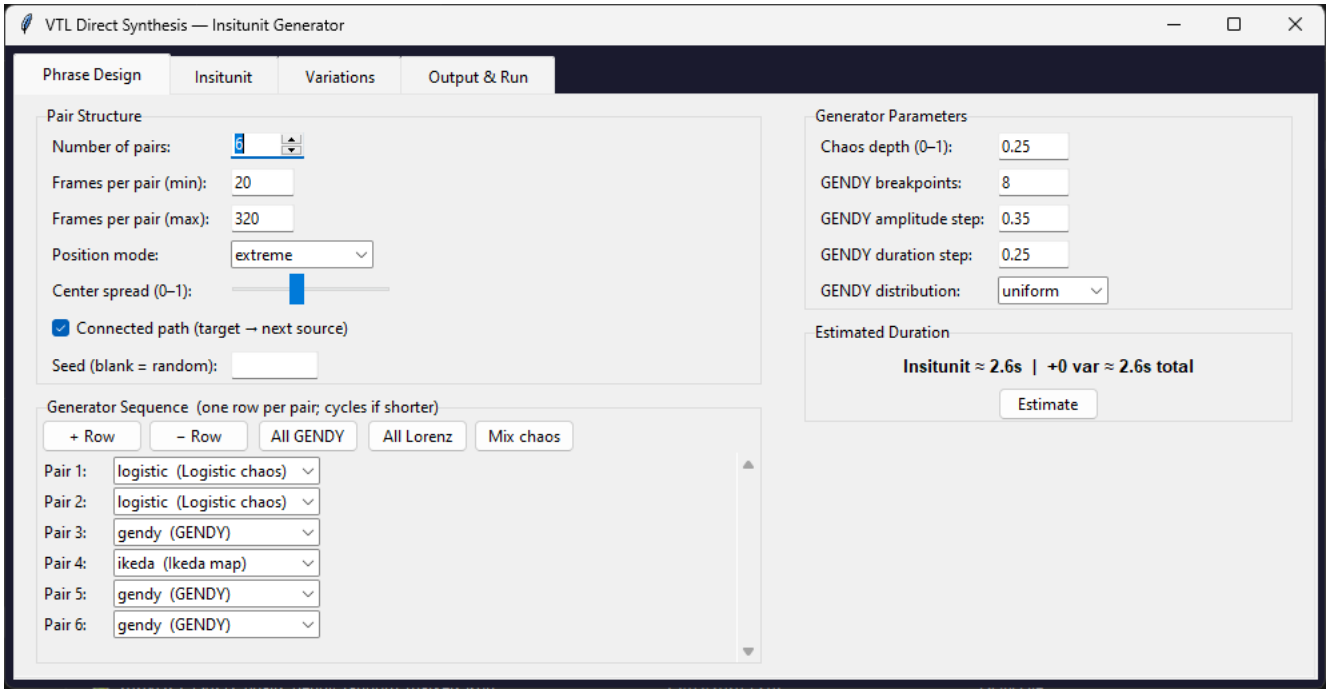


Figure 2.6: (b) Phrase design tab in VIS.

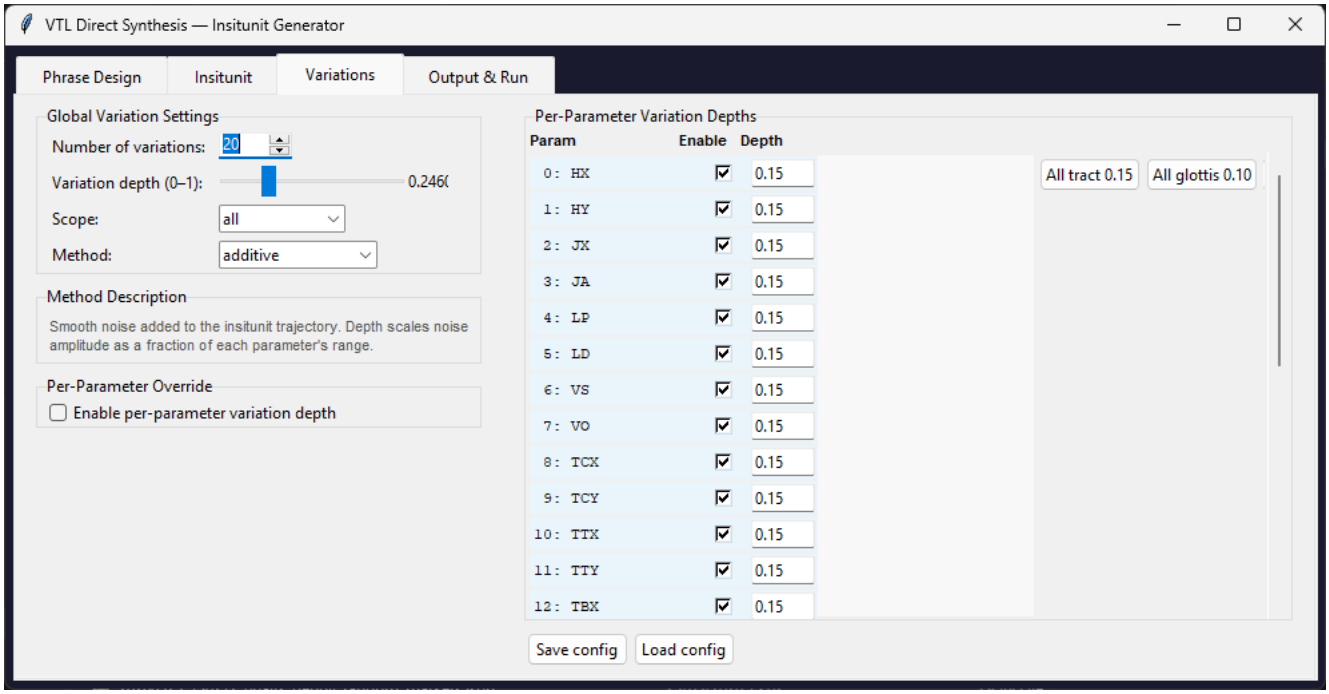
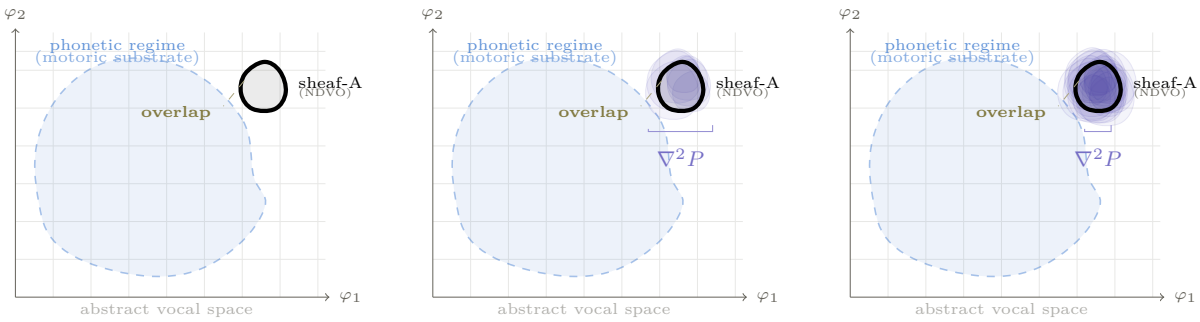


Figure 2.7: (c) Variation tab in VIS.



(i) initial NDVO—*sheaf-A* defined (ii) stalk forming—7 *ALS* (iii) stalk saturated—20 *ALS*

Figure 2.8: Stalk accumulation in abstract vocal space (φ_1, φ_2). Axes carry no numeric scale; proximity encodes phonological relatedness. The dashed region marks the phonetic regime (motoric substrate); *sheaf-A*—the NDVO global section—sits at the edge of the regime, partially overlapping it. The purple shapes are all alternative local sections; the stalk is the thickening presence-by-differential in the space, denoted $\nabla^2 P$.

these gestures. This is entirely different than doing some processing on a voice sample, as in, our intervention respects the causal structure of voice production while simultaneously redirecting it. The motoric substrate mentioned above is retained, but it sheds the *phonological mask*. How these gestures are then manipulated and structured over time follow the logics of insistent modulation, or *iterations of minimally deviating morphological emphasis*:

> Modifications that are small enough in magnitude to preserve the gestural character of the voice but sufficiently targeted to make a phonetic dependency audible by distorting it.

These emphases are the *infrathins* which rub against the linguistic and encoded voice. What do these emphases look like and how are they implemented? This program makes use of `Tensortract` to perform this complex acoustic-to-articulatory inversion. For example, listen to sound excerpt 10 and its resynthesized version 11 containing fifteen iterations of insistent modulations on specific voice shapes. A visual representation of the corresponding movement of the horizontal hyoid position, expressed as HX, over the duration of this audio sample is given in Figure 2.6(a) and in (b) its pitch contour.

(a) (b)

Figure 2.9: (a) Horizontal position of the hyoid over time HX; (b) $f0$ contour

For some gestures, there is less of an intuition how a certain bend or steepness may impact the utterance. For example, the sound created by a fast ascending and descending lip distance LD will labialize the whole phonation as expected. Oppositely, for a similar gesture of the vertical hyoid position HY its spectral profile and dynamics in loudness are harder to predict. Knowing that I will contract and extend the length of the vocal tract sinusoidally only gives confidence that a timbral oscillation will occur, not what the specific timbral change will be.

For an in depth run-through and explanation of this program, navigate to Appendix 3.

2.6.3 In-between NDVO synthesis DVO resynthesis

The example in excerpt 12 merges the two approaches. Slight interventions on the captured time-series by modulations coming from abstract non-linguistic, general functions (e.g., periodic and stochastic behaviors unit generators) have been discussed. Modulations of this kind are labeled as superpositions on time series where an almost complete overlap of the phonetic regime a.k.a motoric subtrate exists. The program can morph two time-series after extracting them via Acoustic-to-Articulatory Inversion of two audio files (Figure 2.10).

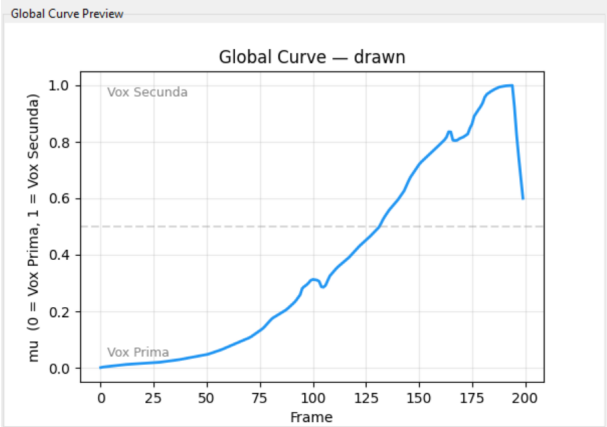


Figure 2.10: Hand-drawn trajectory between $Vprim$ and $Vsec$

In this case, only one audio file is inverted and the other articulatory trajectories are generated by NKOAPP and loaded into the program.[35] The amount of morphing of the first audio file (Vox Prima or $Vprim$) with the other file (Vox Secunda or $Vsec$) is denoted by the morph-value μ and takes on the value between $[0, 1]$ with 0 meaning no morphing leaving us with the unaffected $Vprim$; 0.5 means 50/50 morphing between $Vprim$ and $Vsec$, and 1 returns the unaffected Vox Secunda. The main idea is that the $Vprim$ takes on the DVO and the NDVO is $Vsec$ and that one morphs between them. Their duration can be different and the $Vsec$ trajectories can be stretched to the frame amount of $Vprim$. In the following example, an excerpt of Australian sound poet

[35] Run the `audio2tract_v1_morph2files_drawreplace_GUI.py` file.

Chris Mann's high-rate articulatory gymnastics serves as $Vprim$ (13) and a computed sequence of chaotic vocal movements $Vsec$ (14). Sound excerpt 15 is linearly morphing from $\mu = 0$ to $\mu = 1$; 16 is the logarithmic version.

In addition to global change to the time series, intervening in individual time-sound shapes is also possible. The option to perform iterative modulations can occur on the three levels of insistunits (phonic-morphic-phrasal) depending on how sensitive the audio segmenting threshold value is set. The algorithm used is a `Python` implementation of the Fluid Corpus Manipulation Toolkit's novelty slicing algorithm and turns out to be useful when wanting to segment a waveform based on structural and timbral changes and not just volume, which turns out ideal to catch these *pulmonic turns* as mentioned in 2.6.2. (Tremblay et al. 2022) $VPrim$ is the audio that is analyzed, and once the cut points are retrieved, the amount of alternative local sections is chosen. Instead of the algebraic operations that create the modulations, the modulating value is the morphing value μ which, again, either globally or locally insists on base shapes of $VPrim$; algorithmically or manually deciding its numeric value.

Chapter 3

Four Saints in Three Acts

Four Saints in Three Acts is a piece based on hand-picked text segments from Gertrude Stein's libretto called *Four Saints in Three Acts* and set to music by American composer Virgil Thompson. Its overall rhythmic technique resembles a hocket, largely based on the sound poetry piece *Voicing through Saussure* by the Swiss BarDem duo (Vincent Barras and Jacques Demierre). In it, both players vocalize certain phonemes and syllables as if they were sharing one mouth. *Voicing through Saussure* feels like a polyphonic monologue, but not with the call and response, counterpuntal quality of a motet. Sustaining the phonologic stream of phonemes seems to keep the parts together in flow and is therefore at the base of this piece. *Four Saints in Three Acts* also contains two voices, an acoustic and synthetic one that voice text-segments in succession. *Four Saints in Three Acts* work is a text-sound composition. Text-sound implies a performance of the sounding kind. Texts must be sounded and thus heard to be "read," in contrast to those that must be printed and thus seen to be read. Units of expression in this intermedium are called vocables or "a word regarded as a unit of sounds or letters rather than a unit of meaning." (Kostelanetz 1980, 14) These are the pre-verbal materials that suggest a future sounding and have the capability of being voiced. The vocables in this work are text fragments of verses from Stein's libretto. The vocables across Acts I, II, and III are each of a distinct kind and operate at a different level of identity. Act I takes the phonic as its baseline (i.e., a single NDVO in itself); Act II the morphic (i.e., both NDVO and DVO); Act III the phrasal. On each of these, insistent modulations are applied to one or more parameters, edging the boundaries that sustain the identity, such as its spectral profile, pitch contour, and spectro-temporal rates of change. Text from each act is sourced and then segmented. Depending on which act the text is sourced from, segmentation will happen at a different level (phone-morph-phrase). The acoustic and synthetic voice (respectively AV and SV) ideally succeed one another to work together to

glue the extracted verse materials into an Aria-like line. Besides, the zigzag notation used in this score is also inspired by BarDem's markup in their "score text."[1]

Act I focuses on the phonic tier. For example, "Come one" is phonetically segmented into:

C m o e

o e n

The phones in bold (top) are enunciated by the *AV* and those in gray (bottom) by the *SV*. The elements are now regarded away from their word structure and are treated as raw materials for string concatenations. Each insistent modulation on a particular phonological time-shape, then, has to maintain a morphological continuity.

Act II allows for insistent modulation on the morphic level, allowing for things to happen to the following sentence:

"Was Saint Ignatius able to tell the difference between palms and Eucalyptus trees."

and is fragmented into morphic units for further emphasis by synthetic and natural voice as follows:

W.AA.Z IH.G SH.AH.S B.AH.L T.EH.L D.IH

S.EY.N.T N.EY EY T.UW DH.AH

AH.N.S W.IY.N AH.N.D K.AH T.AH.S

F.ER B.IH.T P.AA.M.Z Y.UW L.IH.P T.R.IY.Z

Each morpheme should again, like in the phonic approach, operate insistently on shapes that just about maintain the overal structure or keep the *sheaf-A* whole. On the phrasal level, such as in Act III, syntactic clauses can also be voiced simultaneously by the *AV* or *SV*, for example:

short longer grass short longer longer shorter yellow grass.

short longer grass short longer longer shorter yellow grass.

pigeons

[1] See https://jacquesdemierre.com/voicing-through-saussure/.

The insitunit in Act I is a phone and, by nature, is non-semantic (NDVO). When the complete acoustic run-through is performed without any hiccups (e.g., wrong accenting, phonological errors such as voicing an unvoiced consonant, too short or too long phonation, etc.), the individual units should chain into an audible English word. And, if the gluing of these chains is strong enough, these phonetic "monomers" can become polymerized into clauses and eventually bond whole phrases together. It should be noted that this difficulty is intentionally there to avoid a strong gluing and to ensure a disruption of the vocal flow. In Act II, the insistunits are regarded more as DVOs. Gluing here is stronger because the *NV* and *SV* both voice clearly demarcated and correctly coarticulated units, therefore avoiding most of the aforementioned phonological errors. For example, the syllables in the word "difference" glue quite easily even when divided among two separate mouths; pronounced as /ˈdɪf.ə.əns/. The *NV* says D IH, the *AV* says F ER and then *NV* again with AH N S. This notation for North American English is borrowed from the Carneggie Mellon University Pronouncing Dictionary[2] where a space visibly separates individual phonemes. The *NV* and *AV* sequentially voice entire clauses in Act III where the denotational hook shows to be well anchored. Across all the acts, the *NV* roams within the phonetic regime. The *SV*, meanwhile, is synthesized so as to push against both the motoric substrate and, especially in the later acts, the boundaries of linguistic associativity. So a sequence of phonic *sheaf-A*'s, if glued strongly enough, can constitute a morphic or phrasal *sheaf-A* at a higher tier. It becomes interesting when an insistent modulation or alternative local section is made on a phonic insistunit (referred to as an *ALS* in 2.6.2). The timbre profile of that phone changes along with the disruption of the coarticulative seam between units and propagates upward and destabilises whatever higher-level *sheaf-A* the sequence was trying to form. The $\nabla^2 P$ rising when iterations of *ALSs* are heard in sequence shows this resistance to venturing towards a linguistic consolidation, i.e., higher-level. Reaching back to Pynchon's concept of the compressed Δt and Stein's prosaic desire to create "a continuous present," the insistent modulation enacts a comparable arrest, both in denotational and non-denotational voiced environments, or what I call *vocales*.[3] The *sheaf-A* is both returned upon and diverged from with each *ALS* making any possible progression insinuated by the to-be-voiced material fight an uphill battle.

[2] See `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`.

[3] Derived from the word *locales* in the architectural sense: "vocale" applies the spatial/locational thinking to voice itself, allowing to hear voice as having an interior topology with addressable sites both in denotational and non-denotational voicings.

Insistent Modulations

The insistunit in each act is treated as a mold from which a host of versions are cast. In other words, an insistunit is the unaltered, non-emphasized vocalization, i.e., *sheaf-A*. Each *ALS* or shape-iteration is the product of an insistent modulation on a specific articulatory trajectory captured by the machine. `NKOAPP` is fed one recording of a natural voicing of the insistunit. Then after analysis in `TensorTract` the acoustic and articulatory movements contained in *sheaf-A* are acquired in a multidimensional matrix, or a tensor. For each insistunit, be it a phone, a phoneme cluster, or a meaningful sequence of words, one or more parameters from 2.1 can undergo insistent modulations, and if heard in sequence, $\nabla^2 P$ only rises. Figure 3.1 shows this signal path for a dummy input. The performer then chooses live how many of these *ALSs* are made audible through speakers.

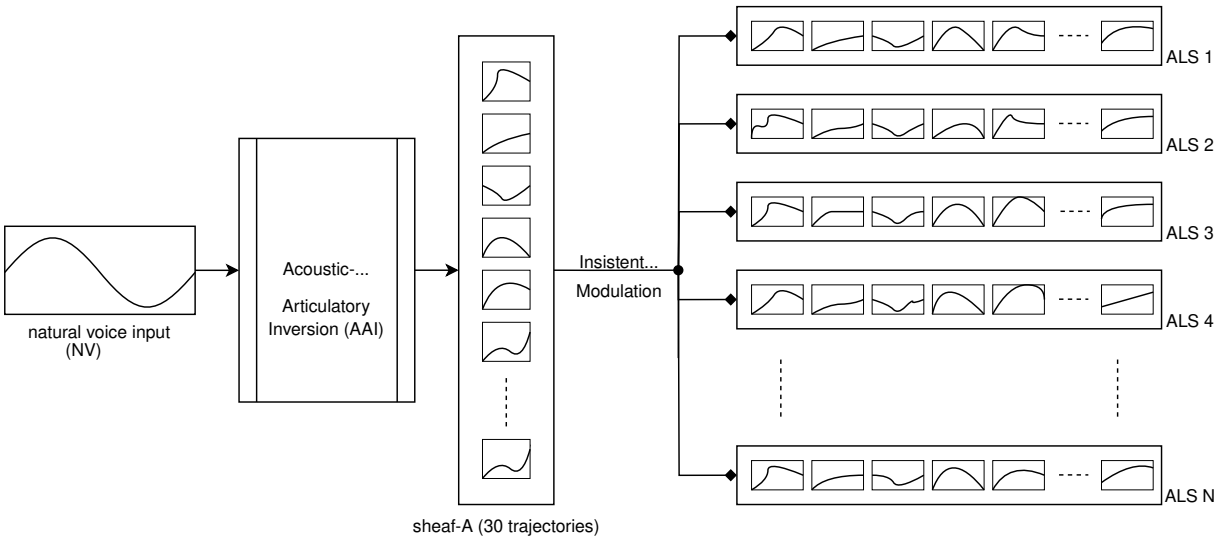


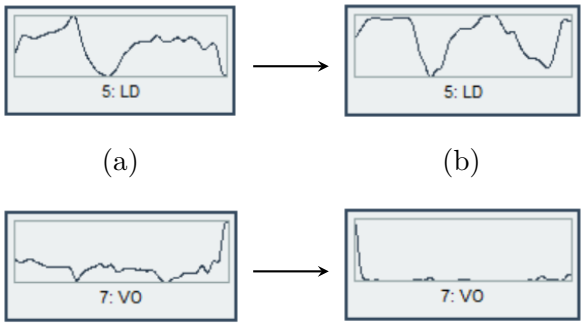
Figure 3.1: Signal flow from *NV* voicing to the Acoustic-to-Articulatory Inversion, and then the subsequent N alternative local sections (*ALSs*)
on *sheaf-A* or insistent modulations on articulatory and acoustic trajectories; in this example all 30 shapes are insisted upon.

The insistunits in the very first phrase of Act I "Come one" undergo nine *ALSs*. In each *ALS*, an insistent modulation occurs as a very small stochastic subtraction and addition to the initial f0 contour. The time-sound shape of this parameter for *sheaf-A* and *ALS1* from the synthetic voicing of "*o*" is given in figure 3.2. For insistunits that are unvoiced such as "*c*," a pitch modulation barely has any audible influence. Eventually these individual sounds are sounded after each other, and necessarily share the same audible canvas with an obvious neighbourhood. It would be shortsighted to address the time-sound shape change of a single insistunit by not considering what it is doing to its neighbours, and what its neighbours are doing to it. This dialectic recalls the question of weak/strong gluing, which, as argued in Act I, is only weakly satisified.

Each shape change demands a bilateral attention to the phrasal sonority, and to the time-sound shapes of the individual insistunit. Later on in part three of this first act, two time-sound shapes are stochastically insisted upon, where this time the resulting shape is more drastically affected. The lip distance LD is being modulated at a rate of $15Hz$ where relatively high values are added/subtracted to its base shape. Parameter number two was the velum opening VO, or, how much of the acoustic energy is transferred into the nasal tract, i.e., how nasal it becomes (sound excerpt 17). This nasal port is fully closed in phonetic neighborhoods of oral stops, oral vowels, fricatives and approximants and now allows for some of the oral sound to be modulated partially into the nasal cavity. Modulation occurs at a rate of $5Hz$, with an amplitude ranging from a minimum of 1% to a maximum of 20% of its full range of motion. The amount of ALS one can choose from here is four, also denoted as $N = 4$. An alternative local section—ALS^1_{LD} and ALS^1_{VO}—for the insistunit "o" is shown next to base trajectories for both LD and VO in figure 3.3.

(a) (b)

Figure 3.2: (a) shows the base f0-contour of *sheaf-A* (also called a *pre-sheave* in Category Theory), and (b) shows the first insistent modulation *ALS1* on this base trajectory.



(a) (b)

(c) (d)

Figure 3.3: (a) and (b) show the base trajectory of respectively LD and VO; (c) and (d) shows one corresponding ALS, denoted by ALS^1_{LD} and ALS^1_{VO}.

Such types of synthetic modulations that embody Gertrude Stein's insistence occur throughout most of Act I, with the last example being the most extreme one out of all. The NV on the other hand tries the same thing but relies heavily on improvised gestures on top of other gestures. The human is the only one that fails in a formal sense, but this is not obvious for the listener as the synthetic voice is actually triggered by the human performer's finger tapping a button, silently embedding ill-timings and stutters into the vocal and performative agency of the SV.

Emphatic repetition on a phrase happens until a satisfactory level of a $\nabla^2 P$ is made, after that, the performer presses a button to move to the next phrase. The full score is given in Appendix 3 and a recording of the full piece can be found in excerpt number 18 .

Conclusion and Future Work

My project set out to find out whether a *vocal presence* can be built through insistent repetition, without having to rely on fixed meaningful units of vocal expression. The answer it arrives at is yes, with conditions.

The first condition is that repetition has to be *insistence* in Gertrude Stein's sense. The revisit to a sound shape must carry some degree of emphasis that makes the iteration distinct from the vocal shape that came before. Essentially saying that deviation in se is the content.

The second condition concerns duration. *Presence-by-differential* ($\nabla^2 P$) is a durational effect that accumulates across iterations instead of arriving after one utterance, building what the thesis calls a *stalk*, a thickening of local presence that becomes audible only after several returns to the same shape. This applies to both *denotational and non-denotational voiced objects*, though the mechanisms differ. In DVOs, the syntactic environment frames the return and NDVOs need a deliberate compositional strategy to provide that framing, due to not having a semantic anchor to bridge an auditory attention between iterations.

The third condition is that the vocal material, whether natural or synthetic, must sit within reach of the listener's *motoric decoding substrate*. A sound is heard as vocal when it carries enough articulatory trace for the listener to recover an implied gesture. Synthetic voice permits manipulation of individual articulatory parameters that remain *coarticulated* and inseparable in natural speech, which made it analytically useful for the compositional work here.

The sheaf framework gave a vocabulary for describing how local vocal identities cohere or fail to cohere across a sequence of insistent iterations. In Castillejo's "TLALAATALA," local sections overlap weakly and almost glue into a global section, producing the halo-like presence the listener is unable to resolve. In Monach's "Automaterga," the clusters are structurally isolated and the only global section holding across the whole grid is the fixed coda, carrying little expressive weight on its own. These two cases represent different relationships between local and global

coherence, and my text-sound composition *Four Saints in Three Acts* moves between them across its three acts.

The triangulation of Homer, Melnick, and Stein established that written texts can solicit vocalization without explicitly instructing it. Homer's meter made vocalization structurally necessary for transmission. Melnick's homophonic translation makes silent reading practically untenable. Stein's iterative syntactic structures push the reader toward phonetic attention, where the sonic shape of the phrase gradually takes over from semantic content. These three cases have shown that a relationship between script and voice depends on how the text handles sound, time, and repetition.

Lastly, the *Four Saints in Three Acts* applies these findings in a setting where a natural voice and a synthetic voice share a single phonological stream, working through Stein's libretto at three levels of granularity, phonic, morphic, and phrasal, with insistent modulations applied throughout on what I termed *insistunits*.

Freudenthal's "Lincos" opened this thesis with a methodological problem: how to design a language (i.e., lingua cosmica) legible to a receiver with no prior knowledge of any natural language. His approach was to build meaning through sufficiently many instances, from which a pattern could be inferred without being formally stated. He called this a *quasi-general definition*. The vocal work in this thesis follows a comparable logic. The *mini-languages* it produces, temporary, formally constrained systems of vocal gesture, accumulate their intelligibility through repeated instances. Most importantly is that these mini-languages are exhausted by the performance that generates them, as the introduction stated, and they accumulate no grammar that outlasts that performance. What Freudenthal's receiver does with mathematical notation, the listener here does with vocal sound, reading more or less present structural relations instead of referential ones.

What I leave open is the question the writing raises most directly. Whether the accumulative quality of presence-by-differential belongs to natural language structures specifically, or whether it holds for any vocal event submitted to sufficient insistence. My music and computer programs only begin to test that claim.

Bibliography

Barthes, Roland. *Camera Lucida: Reflections on Photography*. 2nd printing. New York: Hill / Wang, 1982. ISBN: 978-0-8090-1398-2.

Bergen, Benjamin K. 'The Psychological Reality of Phonaesthemes'. *Language* 80, no. 2 (June 2004): 290–311. ISSN: 1535-0665, visited on 02/06/2025. https://doi.org/10.1353/lan.2004.0056. https://muse.jhu.edu/article/169798.

Bergsland, Andreas. 'Experiencing Voices in Electroacoustic Music'. Doctoral Dissertation, Norwegian University of Science and Technology, 10th May 2010.

Birkholz, Peter. 'Modeling Consonant-Vowel Coarticulation for Articulatory Speech Synthesis'. Ed. by Francesco Pappalardo. *PLoS ONE* 8, no. 4 (16th Apr. 2013): e60603. ISSN: 1932-6203, visited on 12/04/2026. https://doi.org/10.1371/journal.pone.0060603. https://dx.plos.org/10.1371/journal.pone.0060603.

Butler, Judith. *Gender trouble: feminism and the subversion of identity*. New York: Routledge, 1999. ISBN: 978-0-415-92499-3.

Calderón-Garrido, Diego and Josep Gustems. 'Temporal Semiotic Units (TSUs) in Audiovisual Analysis', 117–127. 1st Nov. 2015. ISBN: 978-1-4438-8105-0.

Cámara, Mateo et al. *Optimization Techniques for a Physical Model of Human Vocalisation*, arXiv:2309.14761, 26th Sept. 2023. Visited on 25/05/2026. https://doi.org/10.48550/arXiv.2309.14761. arXiv: 2309.14761[eess.AS]. http://arxiv.org/abs/2309.14761.

Carpenter, Edmund et al. *Eskimo realities*. 1st ed. New York: Holt, Rinehart / Winston, 1973. ISBN: 978-0-03-007576-6.

Castillejo, José Luis. *TLALAATALA: livre d'artiste*, ISBN: 9788460719809 Publisher: Alga Marghen.

Cholin, Joana, Willem J. M. Levelt and Niels O. Schiller. 'Effects of syllable frequency in speech production'. *Cognition* 99, no. 2 (1st Mar. 2006): 205–235. ISSN: 0010-0277, visited on 06/04/2026. https://doi.org/10.1016/j.cognition.2005.01.009. https://www.sciencedirect.com/science/article/pii/S0010027705000600.

Chopin, Henri. *Poésie sonore internationale*. Trajectoires 1. Paris: J.-M. Place, 1979. ISBN: 978-2-85893-032-6.

Deleuze, Gilles, Félix Guattari and Gilles Deleuze. *Mille plateaux*. Capitalisme et schizophrénie / Gilles Deleuze; Félix Guattari 2. Paris: Éd. de Minuit, 1980. ISBN: 978-2-7073-0307-3.

Demierre, Jacques. 'Voicing through Saussure'. Visited on 19/05/2026. https://jacquesdemierre.com/voicing-through-saussure/.

Fadiga, Luciano et al. 'Speech listening specifically modulates the excitability of tongue muscles: a TMS study'. *The European Journal of Neuroscience* 15, no. 2 (Jan. 2002): 399–402. ISSN: 0953-816X. https://doi.org/10.1046/j.0953-816x.2001.01874.x.

Freudenthal, Dr Hans. *Lincos; design of a language for cosmic intercourse.* Studies in logic and the foundations of mathematics. Amsterdam: North-Holland Pub. Co., 1960.

Garcia, Tristan, Mark Allan Ohm and Jon Cogburn. *Form and object: a treatise on things*. Speculative realism. Edinburgh: Edinburgh University Press, 2014. ISBN: 978-0-7486-8149-5 978-0-7486-8150-1 978-0-7486-8151-8 978-0-7486-8152-5.

Gavrilov, A. K. 'Techniques of Reading in Classical Antiquity'. *The Classical Quarterly* 47, no. 1 (1997): 56–73. ISSN: 0009-8388, visited on 06/01/2026. https://www.jstor.org/stable/639597.

Genette, Gérard. *Paratexts: Thresholds of Interpretation*. Literature, Culture, Theory 20. Cambridge ; New York, NY, USA: Cambridge University Press, 1997. ISBN: 978-0-521-41350-3 978-0-521-42406-6.

Grouwels, Joris, Nicolas Jonason and Bob L. T. Sturm. 'Exploring the Expressive Space of an Articulatory Vocal Modal using Quality-Diversity Optimization with Multimodal Embeddings'. In *Proceedings of the Genetic and Evolutionary Computation Conference*, 1362–1370. GECCO '25: Genetic and Evolutionary Computation Conference. NH Malaga Hotel Malaga Spain: ACM, 14th July 2025. ISBN: 9798400714658, visited on 28/11/2025. https://doi.org/10.1145/3712256.3726313. https://dl.acm.org/doi/10.1145/3712256.3726313.

Jakobson, Roman, Morris Halle and Moris Halle. *Fundamentals of Language*. Originally published 2002 as a reprint of the second, revised edition from 1971. Janua Linguarum. Series Minor 1. Berlin;New York: De Gruyter, 2010. ISBN: 978-3-11-017283-6 978-3-11-089426-4. https://doi.org/10.1515/9783110894264.

Kostelanetz, Richard. *Text-Sound Texts*. 1980. Visited on 26/03/2025. https://monoskop.org/images/d/d7/Kostelanetz_Richard_ed_Text_Sound_Texts_1980.pdf.

Krug, Paul K. et al. 'Self-Supervised Solution to the Control Problem of Articulatory Synthesis'. In *INTERSPEECH 2023*, 4329–4333. INTERSPEECH 2023. ISCA, 20th Aug. 2023. Visited

on 09/04/2026. https://doi.org/10.21437/Interspeech.2023-2173. https://www.isca-archive.org/interspeech_2023/krug23_interspeech.html.

Krug, Paul Konstantin et al. 'Precisely Controllable Neural Speech Synthesis'. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. ISSN: 2379-190X, ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Apr. 2025. Visited on 09/04/2026. https://doi.org/10.1109/ICASSP49660.2025.10890772. https://ieeexplore.ieee.org/abstract/document/10890772.

Laing, R. D. *Selected Works of RD Laing: Knots V7*. Selected Works of R d Laing Ser. Florence: Taylor / Francis, 2012. ISBN: 978-0-203-21048-2 978-0-415-19824-0.

Lansdown, John. 'INTERACT: Machine:Man:Society', 27th Aug. 1973. Visited on 13/05/2026. https://computer-arts-society.com/casarchive/cas/uploads/interact-documents-1973.pdf.

Loggins, Vernon. *I Hear America*. New York: Thomas Y. Crowell & Co, 1937.

Mc Guire, Lawrence. *Parse Tree, based on Gertrude Steins poem*. Oscillation Festival 2026 Reader. Brussels: Q-02, 2026. https://oscillation-festival.be/2026/wp/wp-content/uploads/2026/04/reader-combined.pdf.

Melnick, David J. *Men in Aïda*. Brooklyn, NY: punctum books, 2014. ISBN: 978-94-91914-04-1.

Melnick, David J. *Men in Aïda*. punctum books, 31st Dec. 2014. ISBN: 978-94-91914-04-1, visited on 06/01/2026. http://archive.org/details/ebeae9d6-7543-4cd4-9fa9-c39c43ba0d4b.

Michaelis, Kelly et al. 'Motor engagement relates to accurate perception of phonemes and audiovisual words, but not auditory words'. *Communications Biology* 4, no. 1 (25th Jan. 2021): 108. ISSN: 2399-3642, visited on 07/04/2026. https://doi.org/10.1038/s42003-020-01634-5. https://www.nature.com/articles/s42003-020-01634-5.

Myachykov, Andriy et al. 'Random word generation reveals spatial encoding of syllabic word length'. _Eprint: https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1111/bjop.12399, *British Journal of Psychology* 111, no. 2 (2020): 357–368. ISSN: 2044-8295, visited on 06/04/2026. https://doi.org/10.1111/bjop.12399.

Neel, Eric S. 'The Talking Being Listening: Gertrude Stein's Patriarchal Poetry and the Sound of Reading'. *Style* 33, no. 1 (1999): 88–105. ISSN: 00394238, 23746629, visited on 15/01/2026. http://www.jstor.org/stable/10.5325/style.33.1.88.

Niekerk, Daniel R. van et al. 'Simulating vocal learning of spoken language: Beyond imitation'. *Speech Communication* 147 (1st Feb. 2023): 51–62. ISSN: 0167-6393, visited on 09/04/2026.

https://doi.org/10.1016/j.specom.2023.01.003. https://www.sciencedirect.com/science/article/pii/S0167639323000146.

Ong, Walter Jackson. *Orality and literacy: the technologizing of the word*. Ed. first published 2002. New Accents. London: Routledge, 2005. ISBN: 978-0-415-28128-7.

Perloff, Marjorie. *Infrathin: an experiment in micropoetics*. Chicago: University of Chicago Press, 2021. ISBN: 978-0-226-71263-5 978-0-226-79850-9.

Pynchon, Thomas. *The crying of lot 49*. New York: Harper Perennial, 2006. ISBN: 978-0-06-091307-6.

Robson, Ernest M. *Orchestra of the language*. In collab. with George A. Smathers Libraries University of Florida. New York,: T. Yoseloff, 1959. Visited on 31/05/2023. http://archive.org/details/orchestraoflangu00robs.

Rogers, Jack C. et al. 'Discrimination of speech and non-speech sounds following theta-burst stimulation of the motor cortex'. *Frontiers in Psychology* 5 (15th July 2014). ISSN: 1664-1078, visited on 07/04/2026. https://doi.org/10.3389/fpsyg.2014.00754. http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00754/abstract.

Rosiak, Daniel. *Sheaf theory through examples*. Cambridge, Massachusetts: The MIT Press, 2022.

Schrader, Vicki. 'Origins of Silent Reading: Timeline' (2010). https://blogs.ubc.ca/schrader/etec-540/origins-and-influence-of-silent-reading/.

Smalley, Denis. 'Spectromorphology: explaining sound-shapes'. *Organised Sound* 2, no. 2 (Aug. 1997): 107–126. ISSN: 13557718, visited on 26/01/2023. https://doi.org/10.1017/S1355771897009059. http://www.journals.cambridge.org/abstract_S1355771897009059.

Stein, Gertrude. *Lectures in America*. In collab. with Internet Archive. London : Virago, 1988. ISBN: 978-0-86068-991-1, visited on 07/09/2025. http://archive.org/details/lecturesinameric00stei.

– . *Matisse, Picasso, and Gertrude Stein, with two shorter stories*. 1. U.S. ed. Barton: Something Else Press, 1972. ISBN: 978-0-87110-085-6 978-0-87110-086-3.

– . 'Sacred Emily'. In *Geography and Plays*. Boston: Four Seas Co., 1922.

– . *Tender buttons: objects, food, rooms*. Mineola, N.Y: Dover Publications, 1997. ISBN: 978-0-486-29897-9.

Stein, Gertrude and Virgil Thompson. *Four Saints in Three Acts*. 1934.

Südholt, David et al. *Vocal Tract Area Estimation by Gradient Descent*, arXiv:2307.04702, 10th July 2023. Visited on 14/09/2025. https://doi.org/10.48550/arXiv.2307.04702. arXiv: 2307.04702[cs]. http://arxiv.org/abs/2307.04702.

Tenney, James. 'Form in Twentieth-Century Music: (196970)'. In *From Scratch: Writings in Music Theory*, ed. by James Tenney et al. University of Illinois Press, 22nd Sept. 2015. ISBN: 978-0-

252-03872-3, visited on 25/05/2026. https://doi.org/10.5406/illinois/9780252038723.003.0006. https://doi.org/10.5406/illinois/9780252038723.003.0006.

The History of Sound Poetry: An Introduction. In collab. with Charles Amirkhanian. KPFA, 14th Nov. 1976. Visited on 15/01/2026. http://archive.org/details/HistoryOfSoundPoetry.

The Philosophy of Composition. In *Wikipedia*. Page Version ID: 1317332160. 17th Oct. 2025. Visited on 15/01/2026. https://en.wikipedia.org/w/index.php?title=The_Philosophy_of_Composition&oldid=1317332160.

Triple, Canopy. 'The Making of Americans with Triple Canopy', 2013. Visited on 15/01/2026. https://canopycanopycanopy.com/contents/the-making-of-americans-2013.

Van Niekerk, Daniel et al. 'Exploration strategies for articulatory synthesis of complex syllable onsets', 635–639. Proc. Interspeech 2022. 2022. Visited on 09/04/2026. https://doi.org/10.21437/Interspeech.2022-10689. https://www.isca-archive.org/interspeech_2022/vanniekerk22_interspeech.html.

Vasse, Denis. *L'ombilic et la voix: deux enfants en analyse*. Points Série essais 392. Paris: Éd. du Seuil, 1999. ISBN: 978-2-02-037169-8.

Watkins, K.E., A.P. Strafella and T. Paus. 'Seeing and hearing speech excites the motor system involved in speech production'. *Neuropsychologia* 41, no. 8 (Jan. 2003): 989–994. ISSN: 00283932, visited on 07/04/2026. https://doi.org/10.1016/S0028-3932(02)00316-0. https://linkinghub.elsevier.com/retrieve/pii/S0028393202003160.

Weide, Robert L. 'The {CMU} Pronouncing Dictionary', 1993. Visited on 19/05/2026. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

Wishart, Trevor and Simon Emmerson. *On sonic art*. 1st July 1996. ISBN: 978-1-315-07789-5. https://doi.org/10.4324/9781315077895.

Xu, Anqi et al. 'Artificial vocal learning guided by speech recognition: What it may tell us about how children learn to speak'. *Journal of Phonetics* 105 (1st July 2024): 101338. ISSN: 0095-4470, visited on 09/04/2026. https://doi.org/10.1016/j.wocn.2024.101338. https://www.sciencedirect.com/science/article/pii/S0095447024000445.

Appendices

Appendix 1:

VIS: Vocal Insistunit Synthesizer

Navigate to the Github page: `https://github.com/hogobogobogo/VIS`

For Windows machines only!

> **Warning:** Run the GUI script only: `directsynth_insitunit_GUI.py`

Direct synthesis patch for generating vocal tract motor sequences *without* audio input. Instead of analysing speech, the tool builds articulatory movements from scratch using a sequence of source–target gesture pairs, each shaped by an oscillator, chaotic attractor, or GENDY-style stochastic function. The resulting base phrase — called the **insitunit** — is stored in memory and used as the template for any number of controlled variations. All sections are concatenated into a single VTL tract-sequence file.

Features:

Feature	Description
Source–Target Pairs	The phrase is assembled from N gesture pairs. Each pair has a source position, a target position, a frame duration, and a movement generator. In connected mode the target of one pair becomes the source of the next, forming a continuous path through articulatory space.
Movement Generators	12 options shape the path from source to target: smooth (`linear`, `sigmoid`, `ease_in_out`, `sine_arc`, `elastic`), chaotic (`logistic`, `lorenz`, `henon`, `duffing`, `ikeda`), and stochastic (`gendy`, `random_walk`). Chaotic generators modulate around the straight-line trend; depth is controlled by `-chaos-depth`.
GENDY	Adapted from Xenakis's GENDY (1992) stochastic breakpoint synthesis. N breakpoints are initialised randomly; amplitude and duration are updated by a random walk drawn from a Cauchy, Gaussian, or uniform distribution. The result is blended with a linear trend to preserve the source-to-target direction.
Insitunit	The base phrase stored as a $30 \times T$ motor array in memory. Displayed as colour-coded parameter strips in the GUI. Saveable and reloadable as `.npy` for use across sessions.
Variations	Generate M variations of the insitunit via five perturbation methods (see below). Scope can be `all`, `tract`, or `glottis`; individual depths can be overridden per parameter via a JSON file.
Audio Synthesis	Synthesizes the combined output directly via `vocaltractlab-cython` and optionally embeds REAPER iteration markers into the WAV.

Variation Methods:

- `additive` — Smooth interpolated noise added, scaled to a fraction of each parameter's range.
- `scale_deviation` — Each parameter's deviation from its centre is scaled by $1 \pm \text{depth}$. Preserves overall shape but alters magnitude.
- `time_warp` — The time axis is locally compressed or expanded. Sequence is preserved; timing shifts.
- `stochastic` — A stochastic scale envelope multiplies each parameter's deviation from

centre (same spirit as the scaling mode in the segmentation patch).

- `gendy_noise` — GENDY-style stochastic breakpoint noise is added to the trajectory, giving organic, non-repeating texture.

Position Modes:

- `random` — Uniformly random within all parameter bounds.
- `center_biased` — Gaussian distribution around parameter centres; spread is configurable (0–1).
- `extreme` — Biased toward the limits of each parameter range.

CLI Examples:

```
# Minimal: 6 GENDY pairs, no variations, tract-sequence only
python directsynth_insitunit.py

# All Lorenz, 8 pairs, save the insitunit for reuse
python directsynth_insitunit.py \
    --num-pairs 8 \
    --generators lorenz \
    --frames-min 150 --frames-max 600 \
    --save-insitunit phrase_a.npy

# Load saved insitunit, generate 5 additive variations
python directsynth_insitunit.py \
    --load-insitunit phrase_a.npy \
    --num-variations 5 \
    --variation-depth 0.2 \
    --variation-method additive

# Mixed chaos, glottis-only GENDY noise variations, seed fixed
python directsynth_insitunit.py \
    --num-pairs 10 \
    --generators gendy lorenz henon duffing \
    --chaos-depth 0.4 \
    --num-variations 4 \
    --variation-depth 0.3 \
    --variation-scope glottis \
    --variation-method gendy_noise \
    --seed 42

# GENDY with Gaussian distribution + per-parameter override file
python directsynth_insitunit.py \
    --generators gendy \
    --gendy-breakpoints 12 \
    --gendy-amp-step 0.5 \
    --gendy-distribution gaussian \
    --num-variations 3 \
    --per-param-config my_depths.json \
    --embed-markers
```

Per-parameter config JSON format (keys are parameter indices 0–29):

```
{
  "0":  0.30,
  "1":  0.10,
  "19": 0.05,
  "20": 0.20
}
```

Parameters not listed are not varied.

Appendix 2:
`Audio2Tract-Suite`: 3 programs — Instructions and Examples

Navigate to the Github page: `https://github.com/hogobogobogo/audio2tract`

A collection of vocal tract parameter manipulation tools for converting audio to articulatory parameters and applying various transformations focusing on insistent modulations on analysed articulatory trajectories. **For Windows machines only!**

Requirements

Package	Installation
`python-flucoma`	`https://github.com/jamesb93/python-flucoma` (follow instructions there)
`tensortract2`	`pip install tensortract2`
`vocaltractlab-cython`	`pip install vocaltractlab-cython`
`soundfile`	`pip install soundfile`
`scipy`	`pip install scipy` (for cubic interpolation)

Setup

Create the output folder structure from the script directory:

```
audio/output/
```

Patches

*`audio2tract_v1.py`

Simple audio-to-speech with manipulations to VTL timeseries.

Command Line Examples:

```
# Basic conversion (no manipulation)
python audio2tract_v1.py my_audio.wav

# Multiply TCX by 1.5 (increase by 50%)
python audio2tract_v1.py my_audio.wav -m TCX multiply 1.5

# Set f0 to constant 150 Hz
python audio2tract_v1.py my_audio.wav -m f0 set 150

# Add 0.5 to TTY
python audio2tract_v1.py my_audio.wav -m TTY add 0.5

# Smooth the pressure parameter with window size 10
python audio2tract_v1.py my_audio.wav -m pressure smooth 10

# Invert TCX around its mean
python audio2tract_v1.py my_audio.wav -m TCX invert 0

# Multiple manipulations at once
python audio2tract_v1.py my_audio.wav -m TCX multiply 1.5 -m TTY add 0.5 -m f0
    smooth 10

# List available parameters
python audio2tract_v1.py --list-params

# List available operations
python audio2tract_v1.py --list-ops
```

*audio2tract_v1_morph2files_drawreplace.py + GUI

Warning: Run the GUI script only:

```
audio2tract_v1_morph2files_drawreplace_GUI.py
```

Morphing patch between articulator timeseries between two audio files.

Features:

- **Vox Prima** and **Vox Secunda** inputs
- Morphs between timeseries over the whole duration
- Each parameter can follow:
  - The global curve (*Global Settings* tab)
  - Its own local curve (*Tract Parameters* & *Glottis Parameters* tabs)
  - Be replaced by the Vox Secunda timeseries

*audio2tract_v1_scaling_dynamicmovement2_freezepass_individual_shift_

Warning: Run the GUI script only:

`audio2tract_v1_scaling_dynamicmovement2_freezepass_individual_shift_segmentandtimemod_stoch_directSynthesis_GUI.py`

Advanced segmentation and time-stretching patch using FluCoMa novelty detection.

Features:

Feature	Description
Novelty Slicing	Auto-segment audio based on spectral changes (spectrum / mfcc / chroma / pitch / loudness)
Time Stretching	Speed up or slow down segments independently
Stochastic Stretch	Stretch factors evolve across iterations (starts at 1.0×, drifts within range)
Parameter Scaling	Dynamic variation with stochastic curves (passthrough / freeze / stochastic modes)
Parameter Shifting	Remap movements between articulators (numeric or visual drag-and-drop)
Iterations	Generate multiple concatenated variations

Stochastic Stretch Distributions:

- `random_walk` – Small random steps each iteration
- `uniform` – Completely random within range
- `gaussian` – Normal distribution around current value
- `brownian` – Accumulating random motion
- `mean_reverting` – Tends back toward center

CLI Examples:

```
# Basic with segmentation
python audio2tract_v2_segments.py audio.wav --enable-slicing

# With stochastic stretch evolution
python audio2tract_v2_segments.py audio.wav \
    --enable-slicing \
    --stochastic-stretch \
    --stretch-min 0.7 --stretch-max 1.5 \
    --stretch-step-size 0.15 \
    --num-iterations 5

# With parameter shifting
python audio2tract_v2_segments.py audio.wav \
    --enable-shift --tract-shift 3 --glottis-shift -2
```

Parameter Reference

Tract Parameters (0–18):

`HX, HY, JX, JA, LP, LD, VS, VO, TCX, TCY, TTX, TTY, TBX, TBY, TRX, TRY, TS1, TS2, TS3`

Glottis Parameters (19–29):

`f0, pressure, x_bottom, x_top, chink_area, lag, rel_amp,`

`double_pulsing, pulse_skewness, flutter, aspiration_strength`

Output

Generates VTL tract sequence files (`.txt`) in `audio/output/` compatible with **VocalTractLab 2.3** (see `https://vocaltractlab.de/index.php?page=vocaltractlab-download` for the download). This program only works on Windows machines.

Appendix 3:
Text-score *Four Saints in Three Acts*

Listen to *Four Saints in Three Acts* (2026) in the sound excerpt listing numbered 18.

"Four Saints: In three Acts"

Lawrence McGuire

May 20, 2026

Saint Theresa introduction part 1

C m o e
o e n
R p a F r t c .
e e t s A
S i t e e a e t d
a n T r s s a e .
S i t h r s n t u r u d d
a n T e e e o s r o n e .
S i t h r s n t e t d
a n T e e a o a e
S i t e e
a n T r s
S i t e
a n

Saint Therese no visitors/no walls part 2

a e i i a o i o e a l
r n n e n u e w l .
a n r a o o .
i e e a u e
a e i i a o i o i a i s e h w l .
r n n e n u e u o e a l
n o i m e a h y i h m
o y i o h n e o i e .
a n e .
i e a
a n e
i e a
a n e
i e a
a n e e e
i e a a

Dead Nun part 3

S c n v n n n n c n v n n n n c
e a a e o e o e a a e y e y e
n v n t n n c n v n t n n c n v c
a a e o a y e a a e o a y e a a e
n a e o a s .
a h v t s y o

Could be photographed having been dressed like a lady and then they taking out her head changed it to a nun and a nun a saint and a saint so. While I was writing the Four Saints I wanted one always does want the saints to be actually saints before them as well as inside them, I had to see them as well as feel them.

As it happened there is on the Boulevard Raspail a place where they make photographs that have always held my attention. They take a photograph of a young girl dressed in the costume of her ordinary life and little by little in successive photographs they change it into a nun. These photographs are small and the thing takes four or five changes but at the end it is a nun

and this is done for the family when the nun is dead and in memorian. For years I had stood and looked at these when I was walking and finally when I was writing Saint Therese in looking at these photographs I saw how Saint Therese existed from the life of an ordinary young lady to that of the nun, And so everything was actual and I went on writing.

plants and enumeration part 4

h a i t s n f n v o e s
y c n h a d i d i l t .

h r s o p n i s
e b f r a s e

o e t m .
b s e s

S e e o r
c n F u .

T r e p l c .
h e a a a e

F u a i o .
o r w d w

F v a a o t d o .
i e n d p s n

S x p r o .
i a a l r

S v n s a l
e e a h w .

E g t n r o .
i a a b r

N n a e t
i e s a .

T n r t r m n .
e a e i e e

part 5

Planting it green means that it is protected from the wind and they never knew about it. They never knew about it green and they never knew about it she never knew about it they never knew about it they never knew about it she never knew about it.
Planting it green means that it is protected from the sun and from the wind and the sun and They never knew about it and she never knew about it and she never knew about it and they never knew about it. No one surrounded trees as there were none. This makes Saint Ignatius Act Two.

W.AA.Z IH.G SH.AH.S B.AH.L T.EH.L D.IH
S.EY.N.T N.EY EY T.UW DH.AH
AH.N.S W.IY.N AH.N.D K.AH T.AH.S
F.ER B.IH.T P.AA.M.Z Y.UW L.IH.P T.R.IY.Z

S.EY.N.T N.EY M.AY.T R.EH.D
IH.G SH.AH.S B.IY

HH.AW N.IY AA.R IH.N HH.AA.R.D
M.EH N.EY.L.Z DH.EH.R IH.T SH.UW
N.EY.L.Z
AH.N.D
S.IH.L N.EY.L.Z S.IH.L D.AH.Z S.AW.N.D
V.ER AH.N.D V.ER N.AA.T
Y.AH
V.AE.L B.AH.L

IH.N M.AO.R T.UW CH.EY.N.JH.D DH.AH
DH.AH N.IH.NG B.IY F.R.AH.M
N.IH.NG DH.AH N.IH.NG DH.AH N.IH.NG
M.AO.R T.UW M.AO.R IH.N M.AO.R

M.EH M.EH S.EY.N.T.S B.IY T.UW N.IY
N.IY N.IY K.AE.N L.EH.F.T M.EH
N.IY S.IY.N L.EH.F.T M.EH M.EH
M.EH S.EY.N.T.S F.AY.V T.UW N.IY N.IY
S.EY.N.T.S

DH.EY L.EH.F.T M.EH S.EY.N.T.S DH.OW.Z
AA.R T.UW N.IY AH.N.D
DH.IY.Z DH.IY.Z
S.EY.N.T.S S.EY.N.T.S S.EY.N.T.S

K.AE.N N.IY F.IY.L N.IY M.UW AH.N.D
EH W.AH.N EH W.AH.N V.IH.NG
M.UW K.AE.N N.IY F.IY.L N.IY
IH.N V.IH.NG EH W.AH.N EH W.AH.N
AH.N.D M.UW
IH.N V.IH.NG

T.UW B.IH
B.IY L.AY.D

S.IY.N T.UW S.IY.N AH.N.D G.AH.N
EY.T W.EY.T W.AH.N B.IH S.IY.N
T.UW AH.N.D S.IY.N HH.AE L.IY S.IY.N
T.UW T.UW TH.R.IY P.AH B.IY
AH AO.R F.AY.V T.UW AY.V
F.AO.R T.AE.CH.T S.IY.N S.EH.N.T D.ER
S.IH.K.S IH.T S.IY.N V.AH.N T.AE.CH.T
S.IY.N L.EH.T M.IH.K.S S.EH AH
L.EH S.IY.N T.UW
IH V.AH.N EY.T W.EY.T

W.AH.N.S AH
IH.N W.AY.L

W.EH.N W.EH.N DH.EH.N W.EH.N
DH.EH.N DH.EH.N M.EH.N T.EH.N
DH.EH.N T.EH.N DH.EH.N DH.EH.N
W.EH.N W.EH.N DH.EH.N T.EH.N
DH.EH.N W.EH.N S.EY.N.T.S S.EY.N.T.S
T.EH.N DH.EH.N W.EH.N W.EH.N
T.EH.N T.EH.N T.EH.N T.EH.N T.EH.N
T.EH.N T.EH.N T.EH.N T.EH.N

T.EH.N

AY T.R.AY
W.IH.L

with withdrawn.

at that time.

and all.
and all.

might it so.
do and doubling
with it at once left
and right.
left left left
right
left
with what is known.
in time.
in time.

within it
within it within it
as a wedding for them in half of the time.
particularly.
particularly.
call it a day.
call it a day.
with a wide water
with within with withdrawn.

as if
a fourth class

pigeons on the grass
pigeons on the grass
alas.

short longer grass short longer longer shorter yellow grass.
short longer grass short longer longer shorter yellow grass.
pigeons
large pigeons
on the shorter longer yellow grass
alas
pigeons on the grass.

if a magpie in the sky on the sky
can not cry if the pigeon on the grass
alas
can
alas
and to pass the pigeon on the grass
alas
and the magpie in the sky on the sky and to try and to try
alas
on the grass
alas

the pigeon on the grass the pigeon on the grass
and alas.
and alas.

in line
in line
and in in line
and in in line
please say it
first in line.
when it is ordinarily thoughtful and making it be what they were wishing at one time insatia bly and with renounced where where ware and wear wear with them with them and where where wi ll it be as long as long as they might with it with it individually removing left to it when it very well way well and crossed crossed in articulately minding what you do.
might be admired for himself alone.

in a minute.

in a minute
by the time
that it is graciously gratification
and might be with them
to be with them
to be with them
to be
to be windowed.

there are very sweetly many very sweetly René very sweetly there are many very sweetly. Fo undationally marvellously aboundingly illimitably with it as a circumstance. fundamentally a nd saints fundamentally and saints and fundamentally and saints. once in a while and where a nd where around around is as sound and around is a sound and around is a sound and around. a round is a sound around is a sound around is a sound and around. around differing from anoin ted now. now differing from anointed now. Now differing differing. now differing from anoint ed now. now when there is left and with it integrally with it integrally withstood within wi thout with drawn as much as could be withstanding what in might might be so. between thirty five and forty five between forty five and three five as then when they were forty five and thirty five when they were forty five and thirty five when they were then forty five and thi rty five and thirty two and to achieve leave relieve and receive their astonishment. were th ey to be left to do to do as well as they do mean i mean i mean. Left to their in their to t heir to be their to be there all their to be there all there all their time to be there to b e there all their to be all their time there.
there are very sweetly many very sweetly René very sweetly there are many very sweetly. fo undationally marvellously aboundingly illimitably with it as a circumstance. fundamentally a nd saints fundamentally and saints and fundamentally and saints. once in a while and where a nd where around around is as sound and around is a sound and around is a sound and around. a round is a sound around is a sound around is a sound and around. around differing from anoin ted now. now differing from anointed now. now differing differing. now differing from anoint ed now. now when there is left and with it integrally with it integrally withstood within wi thout with drawn as much as could be withstanding what in might might be so. between thirty five and forty five between forty five and three five as then when they were forty five and thirty five when they were forty five and thirty five when they were then forty five and thi rty five and thirty two and to achieve leave relieve and receive their astonishment. Were th ey to be left to do to do as well as they do mean i mean i mean. left to their in their to t heir to be their to be there all their to be there all there all their time to be there to b e there all their to be all their time there.

with wed led
said with led dead said with dead led said with said dead
led wed said wed dead led dead
led said wed.

with be there all their all their time there be there vine there be vine time there be the re time there all their time there.
with be there all their all their time there be there vine there be vine time there be the re time there all their time there.