Distant Listening or Playing Visualisations Pleasantly with the Eyes and Ears

Author: Tanya Clement

Abstract

From Virgil Thomson’s 1934 musical score for Gertrude Stein’s opera *Four Saints in Three Acts*, to readings of *The Making of Americans* (1925) using digital tools, artists and scholars have used musical composition and computational tools to help express something otherwise obscured in traditional readings of Stein’s texts. At first glance, it seems there is little in common between these efforts: readers argue that setting a Steinian text to music is a welcome abstraction that brings the text to new heights of expression, while computing them seems to concretise their subtleties and diminish their expressive possibilities. I argue that it is the situated hermeneutical framework in which visualisations, like musical scores, invite us to "play," that guides our scholarly, interpretive performances with Stein’s texts. In particular, the visualisations I introduce here help us reconsider the significance of oral traditions in Stein’s self-proclaimed magnum opus, *The Making of Americans*.

Modernists have a complicated history with both technology and music. In adjusting to the new technologies of the early part of the twentieth century and in response to an increase in modes of communication, in information, and in the technological transformations of text (as telegraph), images (television and cinema), and sound (telephone and...
Modemists emphasised the role the medium plays in one's experience of art and of the world. Modernist writers were particularly concerned with the extent to which medium played a role in shaping the self as technological innovations and new media comprised, more and more, their means for communicating that self. In particular, music as a medium seemed to provide for an abstraction or expression of the inner self, directly or completely unmediated. As a result, in order to work against this constant state of mediation, Walter Pater claimed, "All art constantly aspires to the condition of music" (Pater 1986, 86).

Gertrude Stein's texts provide an opportunity to question some of these modernist claims about the relationships of music, text, and technologies within the context of today's interpretive landscape of computational technologies. From Virgil Thomson's 1934 musical score for Stein's opera *Four Saints in Three Acts*, to readings of *The Making of Americans* (1925) using digital tools (Clement 2008), artists and scholars alike use musical composition and computational tools to help express something inherent to Stein's texts that was otherwise obscured with more traditional readings. Other Stein texts for which Thomson has composed music includes "The Mother of Us All" for opera, "Susie Asado," "Preciosolla," and "Portrait of F.B." for solo voice and piano, "Portrait of Gertrude Stein" for violin unaccompanied, "Portrait of Alice Toklas" for violin and piano, and Piano Sonata No. 3 on white keys (Thomson et al. 1996, 4). At first glance, it seems there is little in common between these efforts: readers argue that setting a Steinian text to music is a welcome abstraction that brings the text to new heights of expression, while computing them seems to concretise their subtleties and diminish their expressive possibilities. First, this discussion will consider what these two interventions (both musical and computational) have in common and, second, what it means to "play" visualisations as scores that declare the ambiguities and uncertainties of a situated, hermeneutical framework. Finally, this discussion includes an example of such "play" in a comparative reading of several texts by Gertrude Stein as visualised by two tools that measure the sound of text.

**Background: reading Steinian texts with music and with computers**

In the introduction to the libretto published for Stein and Thomson's opera *Four Saints in Three Acts*, Carl Van Vechten wrote that the music was a "perfect complement to the finely singable text which it always enhances and never obscures. The music is as transparent to colour as the finest old stained glass and has no muddy passages" (2008, 6). Brad Bucknell goes further, saying that Thomson's music does more than complement the meaning of Stein's text; it is the space in which meaning making happens: "Indeed, somberness, spirituality, gayness, and so forth, are really, if present at all, made so by the conventional image repertoire created by the music," Bucknell writes. "The language itself will never completely concretise anything, as we have seen" (2001, 211). Kenneth Goldsmith has a similar response to Gregory Laynor's 2008 recording of Stein's *The Making of Americans* (1925) in which Laynor reads and then starts to sing the text on page 135: "It is numbing; it's repetitive; it’s really boring," Goldsmith says of Laynor's initial reading, but then he notes, "and so what happens by page 135 is that Mr. Laynor becomes interpretive and expressive and he begins singing The Making of Americans" (2010). For Goldsmith, the text is inexpressive or "numbing"—it inhibits the production of meaning—but the singing is "expressive" and "interpretive"—singing gives a sense of narrative and knowledge production to the text. As Wendy Salkind says of her theatre students who, confused by a Stein text, began to read it out loud: "they turned it into a square dance or a waltz. They started clapping it out" (Personal interview, July 13, 2011). In each of these cases, music brings out latent meanings in an otherwise inexpressive or unreadable text.

This attempt to engage with the meaning of the text by setting it to music is similar to attempts to read Stein texts with computational analysis. Even before modern-day techniques for computational analysis were widely used, Edith Thacher Hurd, the wife of Clement Hurd, who illustrated many printings of Stein's children's book *The World is Round* (1939), describes this book as a text that is best read within a culture that is enmeshed in advanced technologies and new media. In a companion essay first printed in the 1986 edition, Thatcher Hurd considers the book's limited success after its first and second printings with Young Scott Books in 1939 and in 1966 to its increased popularity.
after its 1986 Arion Press edition. She writes: "The core of meaning in the round songs and the rhyming prose is more comprehensible than it was when the book was first published. Perhaps the electronic age, the age of television and the computer, has enabled us to move along the lines of thought with a speed of cognition that can keep up the swift pace of this expatriate genius." (1988, 158).

More recently, Kenneth Goldsmith describes composer Warren Burt’s 1998 rendition of "Miss Furr and Miss Skeene," which is read aloud by computer voices, as indicative of the "true nature of the structure and the form of Gertrude Stein’s repetitious texts" (2010). Burt, Goldsmith says, has successfully "tak[en] the emotion out of Gertrude Stein’s voice and presentation" (2010). Like Hurd, Goldsmith thinks that Stein’s texts are more understandable within the context of computers since "this was a type of repetition that people weren’t accustomed to in the early part of the century," but "transposed to the computer voices that we’re so accustomed to today, Gertrude Stein’s text makes absolute sense; it’s a sort of emotional flattening, freeing up of the text to become self-sufficient" (2010). While Bucknell heralds the musically inspired versions of Stein’s texts as emotionally heightened, Goldsmith consistently praises computationally composed renditions as more "natural" and "self-sufficient" (i.e., understandable) texts because they are "emotionally flattened." In either case, Bucknell and Goldsmith see musical and computational interventions as key to reading and understanding the texts’ ongoing processes of meaning-making.

Linda Dusman and Wendy Salkind recently composed a performance of Stein’s story Miss Furr and Miss Skeene that exemplifies the process through which a composer can use both computational and musical systems to "sound out" or read the text. In an interview I had with Dusman on July 13, 2011, she admits that they chose Stein’s story because they "fell in love with abstraction" and "to read it out loud suddenly it all made sense -- the rhythm," but Dusman also perceived that there was a system in place in the composition of the text that was making sense to her: "I needed to be true to the text," she says, "Stein is so rhythmical so I used just percussion . . . used dead strokes . . . not flowery." Because she believed that Stein had a compositional system, Dusman’s process of setting Stein’s story to music became a systematic translation in which she transposed the sounds into computational and musical systems. First, she transposed the sounds into numbers for graphing: "I went through the whole thing and listened for the phonemes, the repeating phonemes and then I graphed all of them, numbers of occurrences, each sentence."

Dusman’s narration of her entire process of setting Stein’s text to music is illuminating in terms of understanding how the process of musical annotation and computation share similar underlying methods for scaffolding (setting concrete and abstract textual elements to an external system) in order to express a reading of the text. Dusman describes the graphing (Figure 1) she used and her process:

Red is the 'a' sound, blue is the 'th' sound . . . green is the 'e' sound . . . I did it by sentence . . . I looked at the average for blue, for the 'th' sound across the entire piece . . . it goes further and further away from the average. So there’s a kind of rhythm there for the average number and for the 'a' sound the average is a little higher. But they all come together in the middle . . . which is the longest and most complicated paragraph . . . that big paragraph 12 . . . 'the dark heavy men'. . . [Stein] totally changes at that point, everything changes . . . and then there’s this kind of fight that goes on with 'a' and 'e' get very close here . . . and the 'th' and the 'e' get very close, but eventually the 'e' sound takes over here . . . and that 'e' brightness was why towards the end of the piece they're all only symbol sounds, shiny, shimmery kinds of sounds so that I kind of reflected that transformation so I turned that into a musical score . . . the 'th' sound was a muted tom-tom; the 'ing' sound was a medium gong; flexatone was the word 'pleasant'; and then when she used the word 'voice,' I used a high timpani bend; the 'e' sound is a high piece of metal; and then Furr is a low wood block and Skeene is a high wood block.(Personal interview, July 13, 2011)

When asked if she wanted to make the sounds correspond exactly to the text, Dusman admitted, "I didn’t want to be that obvious about it, and that would have been too cluttered so I did a kind of averaging . . . sometimes it lined up with her voice and sometimes it didn’t line up with her voice" (Personal interview, July 13, 2011). Systematically concretising the text was just one phase in the interpretive process: musical and computational transpositions create
It is in the space of interpretive activities that the concrete and abstract qualities of the textual, the musical, and the computational become so inextricably mixed for Dusman. Dusman, for example, uses a mixture of concrete specifics such as "averaging" the occurrence of sounds and corresponding the occurrence of sounds like "th" or words such as "pleasant" or the characters, Miss Furr and Miss Skeene, to the use of particular instrumental sounds, but she also refers to elements of music that are more abstract when she talks about matching the text to music: "I wanted for it to sometimes be spot on and sometimes just be like an aura . . . there would be kind of like a spatialization or a sound world that would be created for each paragraph so you’d hear the changes from paragraph to paragraph and you’d hear the changes across the course of it but it wouldn’t be really obvious . . . ." (Personal interview, July 13, 2011).

Similarly, Stein wrote that she "took individual words and thought about them until I got their weight and volume complete and put them next to another word;" only then she discovered, "there is no such thing as putting them together without sense" (1990, 504). That "sense" is Dusman's "aura" or the space in which interpretation happens based on complementary concrete and abstract elements within a perceived system. Juxtaposing words can encourage multi-layered meanings while mixing musical notes can inspire epistemologies. Likewise, computation or quantification can foster revelations. Particular examples include John Burrow's classic 1987 text *Computation into Criticism*, in which he argues, "From no other evidence than statistical analysis of the relative frequencies of the very common words, it is possible to differentiate sharply and appropriately among the idiolects of Jane Austen’s characters and even to trace the ways in which an idiolect can develop in the course of a novel" (4) (other examples appear in Ramsay 2011 and Clement 2008). Using digital and musical interventions to interpret texts means both concretising and abstracting as part of the interpretive experience.

**Figure 1: Miss Furr and Miss Skeene visualised by composer Linda Dusman.**

**What computational and musical renditions tell us about the nature of representing literary texts**

Thus far, we have discussed the composer's perspective on using musical and computational interventions for expressing an interpretive reading of the text. Yet, it is at the moment at which the text—whether it has been rendered in paragraphs on the typographical or manuscript page or in musical staffs or as data in bar graphs—is "played" that the reader or listener enters that space of interpretive activity. A sense of "playing" a visualisation is important to note in the context of this discussion because, in the digital humanities, we still tend to "read" a visualisation on a computer screen, whether in tables, in heat maps, or in elaborate and networked spidery nodes. This notion of scholars "reading" visualisations relates to composers or musicians who read scores, which are, in a very real sense, visualised musical compositions. That is, it is usually understood that the musical score is an attempt to represent complex relationships such as the co-occurrence of multiple elements across time and space. As such, it is read, but it is also meant to be played, to be spatialised in time and embodied by voices (or instruments) within a certain physical and hermeneutical context. I am arguing that the same is true of computational visualisations of text. One "reads" a visualisation, but to "play" the visualisation is to engage the spatialised interpretation of that visualisation as an embodied reader in a situated context within a specific hermeneutical framework.

While the readers and scholars discussed earlier often correlate visual representations (images) and aural representations (sounds) with textual interpretations, Stein questions the fluidity our understanding of these interpretive boundaries might suggest. Stein writes, "It is quite natural that some hear more pleasantly with the eyes than the ears," (1993, 90) and elsewhere, she wonders "just what one saw when one looked at anything really looked at anything. Did one see sound, and what was the relationship between colour and sound, did it make itself by description by a word that meant it or did it make itself by a word in itself" (Stein 1988, 191). It is less of leap to
understand the visualisation of sound as a musical score than it is to imagine the kind of synesthesia that Stein describes. On the other hand, Charles Bernstein fluidly constellates the three (image, sound, and text) in an ever-revolving orbit when he advocates for the study of poems as audio instead of print: "Poems set adrift from their visual grounding in alphabetic texts," Bernstein writes, "might begin to resemble the songs from which, for so long, they have been divided" (2011, 115). Further, Roland Barthes introduces the very useful element of performance into this tripartite relationship when he further notes that "there is a progressive movement from the language to the poem, from the poem to the song and from the song to its performance" (1978, 186). This "progressive movement" in Barthes’s workflow is the location in which we discover the hermeneutic framework in which we may learn to understand how we "play" visualisation in digital humanities.

Learning to locate the hermeneutic framework within the interpretive space afforded by a visualisation is seemingly more complex than understanding that the performance of a musical score is the space in which multiple interpretations are remixed. First, in seeming contrast to the move from musical score to performance—from two-dimensional representation to four-dimensional life in which we understand that there are multiple levels of interpretation such as the composer’s, the producers’, the performers’ and the distant listener’s—computing seems to distill the many-layered four-dimensional space of the text in performance (i.e., embodied within the performance network of interpretations with the listener in time and space) into a two-dimensional script called "code." In other words, historically, the musical score is understood as an attempt to represent complex relationships such as the co-occurrence of multiple elements across time and space, a representation that is a provocation to further interpretation. However, we often flatten computational visualisations as reductive, positivist representations of "truth." Johanna Drucker suggests how computational visualisations can afford this same interpretive space. She argues that the "task of representing ambiguity and uncertainty has to be distinguished from a second task – that of using interpretations that arise in observer–codependence, characterised by ambiguity and uncertainty, as the basis on which a representation is constructed" (2011). Likewise, instead of reading a two-dimensional visualisation as the inevitable result (or complete representation) of data analysis, we learn to read visualisations as musical scores, as signposts pointing towards many possible interpretive "results" or readings. This is to say that visualisations of computational data are as inexact a representation of a text or of many texts as musical scores are inexact representations of an operatic performance. Visualisations provide an algorithm for an interpretive performance that we necessarily understand is embodied and situated and multilayered. As such, we should learn to read visualisations as a means for starting to read, as a script that provides directions like the musical score that presents as hermeneutical guide posts a measure, a key, and time signatures.

This approach to reading visualisations as hermeneutical guides is reinforced by Drucker’s call for us to understand "data" (or the "given") as "capta" (or the "taken") in digital humanities. An unproblematised understanding of capta, Drucker argues, lead us to perceive visualisations as "fact" (2011). A related phenomenon occurs in the modernist move toward music. In describing the issues the modernists faced, Bucknell inadvertently sheds some light on why we still do not know how to use computational visualisations in our interpretive activities. Bucknell writes:

\[\text{... the very notion of the inward, of subjective consciousness and the problem of what might constitute a representation of it are not simple things to separate ... music is therefore not simply linked as a metaphorical solution to the representation of consciousness, rather it becomes a part of the larger question of the problem of representation in modernism itself. (2001, 3)}\]

Likewise, the sense that a visualisation is two-dimensional or a simplification of textual complexity is not the problem. The problem is as Bucknell describes: it is the problem of representation itself, but I would add that in the case of reading visualisations in digital humanities, the problem of representation can be alleviated by a shift in expectations. That is, the reader reads the musical score with the necessary perspective of capta rather than data.
Audacity

Considering the efficacy (how well the visualisations allow for interpretation) of two different kinds of visualization tools helps demonstrate the important role that an awareness of Drucker’s "observer-codependence" can play. The first tool is Audacity, a free, open-source tool that allows a reader to create waveforms and spectrograms with audio files. What I have visualised in the first example (see Figure 2) is three waveforms of three readings (one per line) of the same three sections of Gertrude Stein’s *The Making of Americans*. I created the first reading represented in the first line using OpenMary (Modular Architecture for Research on speech sYnthesis), an open-source text-to-speech system that "reads" texts (creates audio files) with a computer-generated, female-gendered, American dialect. The second reading, in line two, is produced from a recording by Gertrude Stein from 1934. The third reading in line three is by Gregory Laynor who, as mentioned previously, created his reading of *The Making of Americans* in 2008. At first glance, this is an interesting visualisation, because the image shows a change (represented by the vertical line in the centre of each reading) at the point at which there is also a change in the nature of the text—it is a break between a more traditional narrative section in which Stein is telling the story of a man and his son and their discovery that pinning butterflies is cruel, and another section in which Stein uses repetition heavily to create an abstract montage of meanings. Here are two representative sentences from **Part A** (from Chapter 5) and **Part B** (from Chapter 9):

**Part A:**

> One of such of these kind of them had a little boy and this one, the little son wanted to make a collection of butterflies and beetles and it was all exciting to him and it was all arranged then and then the father said to the son you are certain this is not a cruel thing that you are wanting to be doing, killing things to make collections of them, and the son was very disturbed then and they talked about it together the two of them and more and more they talked about it then and then at last the boy was convinced it was a cruel thing and he said he would not do it and his father said the little boy was a noble boy to give up pleasure when it was a cruel one. (**2008**, 489)

**Part B:**

> Any family living going on existing is going on and every one can come to be a dead one and there are then not any more living in that family living and that family is not then existing if there are not then any more having come to be living. Any family living is existing if there are some more being living when very many have come to be dead ones. (**2008**, 925)

One reading of Figure 2 is that this is a visualisation of Goldsmith’s point that computer voices bring Stein’s "tropes of repetition to a computer inspired level" of intensity. Indeed, in this figure, we see that the computer voice is seemingly more dynamic—there are higher peaks and lower valleys—but when it comes to the repetitious section the human voices (of Stein and Laynor) seem flattened in comparison to the previous part. While this is an exciting hypothesis, the visualisation is misleading: a waveform simply identifies volume and tempo—the computer voice is speaking more loudly and at a faster rate. It is possible that comparing volume and tempo is of interest to a user, in which case, comparing files in the waveform view would require establishing a common benchmark among the files. The light and dark blue sections of the visualization, as well as the vertical scale on the left can help establish this benchmark. The dark blue pixels represent the highest values or the loudest sample in the group, while the light blue area is the average Root Mean Square value or average sound level for the same group of samples. The vertical scale measures the amplitude or level or magnitude of a signal. Bernstein makes the claim that waveforms can only
identify part of what makes poetry audio files interesting. "There are four features or vocal gestures, that are
available on tape but not page that are of special significance for poetry," Bernstein writes. These include: "the cluster
of rhythm and tempo (including word duration), the cluster of pitch and intonation (including amplitude), timbre,
and accent" (2011, 126). If we agree with Bernstein, that these clusters are what signify meaning in sound, then it
would be a stretch to argue that the waveforms, which only visualise part of the first cluster (tempo) and part of the
second cluster (amplitude), visualise the meaningful features to which Goldsmith vaguely refers when he calls Burt’s
computer readings "inspired." This is to say, while the image is showing us patterns, the patterns are a measurement
of volume and tempo intensity, which are features of sound that do not signify much to us within the hermeneutic
framework of "special significance for poetry" that Bernstein has identified.

The second visualisation (Figure 3) is a spectrogram created within Audacity of the same readings; Figure 4 is a
zoomed view of the same spectrogram on the words "... some such thing. Family living ..." Unlike a waveform, a
spectrogram shows the information necessary to plot prosody features that include timbre and accent, features to
which Bernstein and others have attributed meaning-making properties. Spectograms show the amount of energy in
different frequency bands over time. In Figure 3 and Figure 4, the blue colour is the least energy and the red and white
are the most with the higher frequencies at the top of the scale and the lower frequencies at the bottom. For example,
in Figure 3 and Figure 4, the spectrograms show the subtle differences that close reading phrases (or "close
listening" in Bernstein’s parlance) has always shown, such as how loudness or amplification changes or corresponds
to different frequencies or kinds of sounds such as consonants and vowels. For example, one can see in Figure 4 that
the consonants and vowels look completely different. The consonants are red floating clouds that have higher
frequencies, while the vowels are bright white spots with lower frequencies. Like Stein, the reader imagines with
these visualisations that she can suddenly "hear more pleasantly with the eyes than the ears" (1993, 90) and that she
can discern a relationship between the patterns she sees and the patterns she hears.

Yet, while Audacity is a powerful tool that can be used to visualise meaningful textual features, the visualisations in
Figure 2, Figure 3, and Figure 4 do not reflect the rhetoric of the established hermeneutic framework. In other words,
the Audacity visualisations are like a musical score without measure and without a perceived audience rather than a
score with a key and a time signature set squarely within the context of a larger musical conversation. The
visualisation is difficult to "play" for meaning because it is flattened of the contextual subtleties a legible
hermeneutic framework brings.

In contrast, in the digital humanities, we must develop tools that articulate the particular hermeneutic and
interpretive activities with which they engage. Consequently, the views present features that are "given," consciously
produced as a specific means for reading according to a particular understanding of how interpretive activity
happens. Like a piano concerto, a particular interpretive experience with a digital tool or visualisation may not always
be generalisable to any audience, for any purpose, with any instrument. Like composing and producing a musical
score, developing a tool in the digital humanities means choosing modes of analysis with particular textual features
whether they be parts-of-speech or sentimental phrases) and modes of visualisations (such as bar graphs or wordles, or waveforms or spectrograms, for example) based on methods and theories that we posit are useful and productive within our community.

ProseVis

The second tool under consideration here is ProseVis, a tool I helped develop to analyse aural features of text (this tool is available for download at http://tclement.ischool.utexas.edu/ProseVis. See Clement et al. 2012 for a description of the development of ProseVis). Based on research and theoretical models within a particular set of hermeneutics surrounding a concept of aurality, ProseVis visualises the pre-speech potential of sound as it is signified within the structure and syntax of text. In Charles Bernstein’s theory of "aurality," he defines the term as the "sounding of the writing," in contrast to "orality," which has an "emphasis on breath, voice, and speech" (1998, 13). ProseVis uses the OpenMary system to create a sound surrogate to represent the aurality of sound (or the pre-speech potential of sound as it is signified within the syntax of a text) in terms of a "best guess" system. Dwight Bolinger helps us define this system with his research in intonation: "in the total absence of all phonological and visual cues, the psychological tendency to impose an accent is so strong that it will be done as a ‘best guess’ from the syntax" (1986, 17). In other words, when we encounter a written word within a phrase, we make "best guesses" for how to read that text out loud based on how that word is used within its syntactical context. The OpenMary XML output represents potential sounds since the utterance never happens. That is, ProseVis does not visualise information from an audio file (as I used Audacity to do in Figure 2, Figure 3, and Figure 4); instead, it visualises the XML transcription that OpenMary produces in the process of creating an audio file. This XML represents a script for OpenMary’s "best guess" for how the text should be read out loud. The XML script is based on the structure of the text, including each word’s part-of-speech, its stress and tone, and its location within a phrase. OpenMary reflects the expertise of its developers—Das Deutsche Forschungszentrum für Künstliche Intelligenz (German Research Centre for Artificial Intelligence), Language Technology Lab, and the Institute of Phonetics at Saarland University—who use the Carnegie Mellon University (CMU) Pronouncing Dictionary as well as a folksonomic technique for representing words that are not in the CMU lexicon. This methodology involves generating a lexicon of known pronunciations from the most common words in Wikipedia or by allowing developers to enter new words manually ("Adding support for a new language to MARY TTS" n.d.). Most importantly, this "best guess" system means the OpenMary output represents the ambiguity of "observer codependence" —it is the result of the researcher-developer’s modelling of a reader’s "best guess" system making meaningful sounds while reading text.

We have developed an interface in ProseVis that allows the reader to interrogate these ambiguities by mapping the features extracted from OpenMary to the words in context (ProseVis was developed in a two-stage process, first as VerseVis: Visualizing Spoken Language Features in Text by graduate students Christine Lu, Leslie Milton, and Austin Myers as part of a graduate course in visualisation with Ben Schneiderman at the University of Maryland, College Park. Megan Monroe further developed the prototype as ProseVis under the auspices of the "Seasr Services" project funded by the Andrew W. Mellon Foundation). Recreating the context of the page not only allows for the simultaneous consideration of multiple representations of knowledge or readings (since every reader’s perspective on the context will be different) but the reader’s ability to change the views also allows for a more transparent view of the underlying information driving the visualisation within the context of a specific hermeneutic framework. For instance, in the right panel shown in Figure 5 and Figure 6, a reader can choose to see "full," "vowel," "beginning," or "end" sounds, to see lines broken by phrase, sentence, or paragraph, and to see the words in context as parts-of-speech or as phonetic spellings. These choices are based on research that indicates that mapping features pulled out through text analysis back onto the text in its original form allows for the kind of reading in which literary scholars ordinarily engage: that is, we read words in the context of phrases, sentences, lines, stanzas, and paragraphs (Clement 2008; Soderstrom et al. 2003). Figure 5 and Figure 6 show patterns of vowels that occur within the context of phrases further contextualised within the larger text. In comparison, the sounds represented in the first lines of
Figure 3 and Figure 4 are from the sound file OpenMary creates from reading the XML script it has produced "out loud." In the below zoomed example (Figure 5 and Figure 6) we see the same line ("... some such thing") visualised using information taken from the XML itself. The excerpt from Figure 5 shows the accent features for each word visualised as peaks and valleys. Like the spectrograms from Figure 4, the excerpt from Figure 5 shows a peak on "th" and "ing," meaning both are emphasised. There is a straight line on "some," which is neutral. The excerpt from Figure 6 shows the words coloured according to each word's verb sounds and we can see new patterns such as the fact that "some" and "such" have similar vowel sounds as "come" and "one" – all of which are shaded the same colour green. At this phase in development, colours are assigned automatically by default. Each unique element is added to a list in the order that it appears in the input file. When the entire input file has been read, the colour spectrum is divided by the number of unique elements, resulting in an equal-sized list of sequential colours. Colours are then assigned to the corresponding entry in the element list. Users can also assign colours to attributes.

This information becomes more interesting as we start to detect larger patterns across the text. Zooming out in the ProseVis interface allows the reader to see the patterns across a larger swath of the text. Scrolling allows the reader to see the patterns in progression across (or in this case, down) the text. The understanding that the underlying tabular information is "capta" coupled with the opportunity to toggle back and forth among the various textual features provides the interpretive space in which scholars can consider all the manners in which this text makes meaning with sound within this hermeneutic framework.

ProseVis has been developed under that assumption that text makes meaning with sound, that syntactical components provide for the "script" upon which sound makes meaning, and that the context of the text is essential for understanding how meaning is made. It defines sound within a particular hermeneutic framework much like a composer chooses the key, the length or pace of a sound, its relationship to other sounds, and its amplification in response to her understanding of what a musical movement should portray and the context of the audience. As a result, the ProseVis interface encourages interpretive activity in which the reader engages in a performance of all of these elements, enacted within the same kind of multilayered context of "observer-codependence" that an entourage of composers, producers, performers, and audience members would bring to a real time performance.

What reading Stein's texts with digital tools can tell us about digital humanities scholarship

Ultimately, visualisations are not the end product, nor is the data where the inquiry begins. The traditions reflected in a computational visualisation are always part of a hermeneutic framework that is also always part of a history of technological and methodological remediations. Other tools for looking at prosody and sound with computational tools include Marc Plamondon’s AnalysePoems and Smolinsky and Sokoloff’s Pattern-Finder (see Plamondon 2006 and Smolinsky and Sokoloff 2006). This consideration of sound in The Making of Americans, for example, contributes to the longstanding inquiry into how oral traditions help us read literary texts.
As a circular text (Clement 2008), *The Making of Americans* draws on oral traditions that incorporated circular compositions. *Making* is not circular in the sense that James Joyce's *Finnegan's Wake* is circular—the ending does not "link" to the beginning—nor is it circular in the same way that Joyce’s *Ulysses*—in which the movement progresses around the clock—is circular. In some ways, the text circumambulates in modes and for reasons similar to those in contemporary feminist literary traditions. In Gayle Green’s text *Changing the Story*, she includes a synopsis of books with a "structure of circular return" that take on feminist critical perspectives, such as Margaret Drabble’s *The Waterfall* (1969), Margaret Laurence’s *The Diviners* (1974), Gail Godwin’s *The Odd Woman* (1974), various novels by Margaret Atwood, Octavia Butler, and Ursula K. LeGuin (1991, 14-17). At the same time, the circular structures harken to an older storytelling tradition that scholars have identified as "ring composition." Mary Douglas’ book *Thinking in Circles* (2007) outlines the history of a long-used literary form that has fallen into disuse in contemporary society, so much so that the first sure sign of a possible ring form, she contends, is the confused reader. Douglas maintains that the repetitious nature of confusing texts like the psalms in the bible, *The Iliad*, or *Tristam Shandy* indicates a ring composition structure (2007, 22). Though many scholars argue that the *Iliad* is structured in this fashion (see Stanley 1993, 307–8; Gaisser and Thalmann among others), scholars such as Stephen A. Nimis assert that "the symmetry does not serve to focus attention on a central element, but is a secondary effect of the way the flow of the discourse is interrupted by the decision to restate or reemphasize something" (1998, 70). Thus, Nimis argues that the structure is the result of the process of composition.

To be sure, the combinatorial aspects of parallelism, including chiasmus, have a long tradition in canonical literature and in ring compositions. James W. Watts argues that the complex characterisation of King David in his psalm called "David’s Thanksgiving" from 2 Samuel 22 is the result of the psalm’s chiastic structure. This part (in which David praises Yahweh), mirrors the chiastic structure of 2 Samuel 21–24 as a whole and serves to emphasise David’s devotion to God even as he goes to war with God’s other children (Watts 1992, 114–117). Robert Alter cites the same "victory psalm" as a good example of such parallelism because it includes "quasi–narrative elements and discrete segments with formally marked transitions" that demonstrate the importance of variation (in stresses, themes, and syntax) within the repetition (Alter 1990, 613). In each of these cases, as in *The Making of Americans*, thematic and syntactic variations contribute to developing complexities of meaning. Needless to say, much study on the bible has been done regarding its parallel structures. Alter writes extensively of the many studies completed on this topic in biblical study. See too Welch and McKinlay’s *Chiasmus Bibliography*, which covers studies in chiasmus across many literary genres and time periods.

Learning to read *Making* as a ring composition is tantamount to understanding the method for making meaning that the text itself lays out for the reader, namely to read the whole in terms of its parts. This view of the text also establishes the hermeneutic framework we have chosen for "playing" the text in ProseVis. For example, Figures 7–10 show another view in ProseVis in which predictive modeling results are highlighted with colours corresponding to one of eight books (or fewer if a book has been "deselected" in the ProseVis interface). Five of these books are by Gertrude Stein, including two shorter prose–poem pieces "Picasso" (1912) and "Matisse" (1912), a longer prose poem *Tender Buttons* (1914), and two novels *Three Lives* (1909) and *The Making of Americans* (1923). The other texts are *The Iliad*, translated by Andrew Lang, Walter Leaf, and Ernest Myers (1882); *The Odyssey*, translated by S.H. Butcher and Andrew Lang (1882); and *The New England Cook Book* (1905). The texts by Stein were works that were written during the same time period as *Tender Buttons*. The works by Homer and Joyce’s *Ulysses* were chosen based on preliminary work I have done comparing the repetition patterns in *The Making of Americans* to these texts (Clement 2008; Clement et al. 2012). The translations of the *Iliad* and the *Odyssey* represent those editions that scholars have identified in both Joyce (Schork 122) and Stein’s (Watson 20) libraries. A description of the algorithm used to create these comparisons is discussed elsewhere (Clement et al. 2012). The texts being compared to these books are the same two parts of *The Making of Americans* from Chapters 5 (Part A) and 9 (Part B). For this analysis we are only using features for comparison that research has shown reflect prosody, such as part–of–speech, accent, stress, and semantic location.
Basic patterns exposed in these figures are indicated by the colours that represent each book. For example, there are differences between the prosodic patterns in Part A (Figure 7), which is a more traditionally constructed narrative, and those in Part B (Figure 8), which is more repetitive and experimental. The colours present in Figure 7 show us the phrases that have prosodic features that are most like one of the other texts in the set. And we can see that according to the algorithm we used, Part A has phrases with prosodic patterns that are very similar to the prosodic patterns in the other Stein texts (with the exception of Tender Buttons) as well as those in the Odyssey and the Iliad.

Figure 9: Part A excerpt from Chapter 5 in The Making of Americans showing comparisons showing vowel sounds.

Figure 8 (which shows text from Part B in Chapter 9) shows us that, according to this algorithm, very few texts in our study have prosodic patterns that correspond to this section, though the portraits of "Matisse" and "Picasso" have the most similar patterns of any of the texts in comparison to this section.

Figure 10: Part B excerpt from The Making of Americans showing comparisons sin ProseVis.

In Figure 9 and Figure 10, in which we have zoomed out to look at the text with each line representing a paragraph, one gets a view of larger patterns or tectonic shifts in the text, which the algorithm has marked as more like "Three Lives" (in red) or more like "Picasso" (in blue).

Figure 11: Excerpt from The Making of Americans showing comparisons in ProseVis.

Figure 12: Excerpt from The Making of Americans showing comparisons in ProseVis.

Using tools to consider what it means to "listen" to Stein’s texts (she who admits "I don't hear a language, I hear tones of voices and rhythms," [1993, 70]) is unsurprising when so many have attempted to "hear" this text in so many different manners. These listenings include an operatic score written by Leon Katz with composer Al Carmines (published in 1973 by Something Else Press) and an annual non-stop 24-hour New Year's Day reading of the text by prominent authors and poets in a New York gallery. These readings were initiated in the seventies at the Paula Cooper Gallery (http://www.artnet.com/paulacooper.html). According to a posting on the Poetics Archive (http://listserv.buffalo.edu/cgi-bin/wa?A0=POETICS), Dec 1997, was the eighteenth such reading and would probably take fifty hours to complete. Much like the opportunities for interpretive activities these performances afford, ProseVis provides for an investigatory environment that allows us to consider possible correspondences between patterns we might hear and the role oral traditions play in Stein’s self-proclaimed magnum opus.

For example, playing this text in this way illuminates three different ties between its composition and the compositional traditions of oral culture that allow us to read the text differently. First, Stein is writing this text within a mythic framework. The creation of family histories has a long tradition in oral storytelling. Ring compositions have figured heavily in classical oral traditions used in both Western and ancient Chinese celebrations for reciting origin myths and for recalling the history of a race, a nation, or—one could say—a family's progress (Douglas 2007, 27). That Stein had this kind of mythic structure in mind is perhaps also indicated by the fact that she began writing the novel by writing its opening parable derived from Aristotles’s Nichomachean Ethics (this attribution is quoted in many places. See Wald 1995 and Mitrano 2005). In addition, circular composition was often the structure of creation myths used "at the beginning of a new political period to affirm the relations between different shrines of the community" (Douglas 2007, 103).
Second, Stein was emphasizing the shifting nature of language as both determined and indeterminate. Using ring composition, Stein was evoking the indeterminacy that results from multiple iterations of the same word used differently and the constrained meanings that patterned cross-referencing provides by ultimately limiting the multiple meanings of words, phrases, and passages to the couplings indicated in the text — providing for that simultaneous experience in language that is at once chaotic and restrictive. Gayle Greene calls the phenomenon of "circular return . . . a brilliant response" for what otherwise could be considered "inaccessible and esoteric" in feminist texts, "since it leaves intact the linear sequence of language and narrative, retaining its coherence and comprehensibility while also critiquing its limitations and suggesting alternatives" (1991, 15).

Third, ring composition allows Stein to use the compositional structure as a meaning-making element that works to weave Stein's language experiments with repetition into the greater goal of the text (Clement 2008). In view of this tradition, The Making of Americans is not only about the progress of an immigrant family in a new land (the United States), but it is about the process of Stein's creating that history in and of itself, since, as oral traditions faded, the ring composition — its complexity requiring a well-trained command of language — became a stamp of authority (Douglas 2007, 30). We see this command of the language in the repeated patterns of sounds (represented by the colours) that we see in Figure 11 and Figure 12. These images corroborate the fact that sound and prosody are also making meaning in the text in similar cyclical patterns. In this way, the material and formal complexity of a ring composition reflects her assertion that she is creating a complete history of every one (or each one).

**Conclusion**

One last example helps illuminate why Drucker's notion of "observer-codependence" is essential in thinking about the work we are doing with visualisation in the digital humanities. In the course of her examination of Lucy Church Amiably, Carolynn Van Dyke observes that working with Stein's text in the context of computational experiments provokes Van Dyke to see and to articulate Stein's process of meaning making in computational terms. She calls Stein's use of various semiotic systems "coding," arguing that "the design of Lucy Church Amiably is the succession of coding rules" or "semiotic systems" focused on [1] "the benevolent deconstruction of bourgeois narrative" (1993, 186); [2] "nonsentences" that are associative and rhythmical (1993, 187); and [3] the "pastoral" connotations (1993, 188). Because the computer's rules and systems for creating what reads like nonsense can be described, Van Dyke argues, one begins to think about how Stein's nonsensical text is also rule-based, even though "her store of data and what might be called her data structures are much richer" and "her algorithms, that is her methods for choosing and combining literary and linguistic structures — can be more flexible, more responsive to previous output" (1993, 182). Van Dyke's method for creating systems of reading is a kind of "distant reading" that allows her to see the chaotic "noise" of the text as sense-making information: "Certain kinds of noise reoccur throughout the novel, each readable once we have identified its significant patterns and its particular kind of referent; and not only the types of significant error, but also their sequence, bear meaning" (1993, 186; emphasis added). The opportunity to investigate significant patterns within the context of a system that can translate "noise" (or seemingly unintelligible information) into patterns, within a hermeneutic framework that allows for different ways of making meaning with sound, opens spaces for interpretation. This work allows for distant listening.

As a culture, we use sound to make meaning, but we have sublimated its study because, like studying "noise," studying sound is inconvenient, logistically burdensome in terms of the richness of the resulting patterns, and requires a daunting degree of agency from the interpreter, who must establish new methodologies for interpretation. Van Dyke admits that her attempt to make Stein's text "readable" is based on the reader's desire "to transform apparent noise into information" (1993, 179), but if we read Van Dyke's "noise" as sound — we realise that this desire is not in contrast to but a complement to more traditional interpretive practices. That is, Van Dyke realises the immense responsibility that these kinds of textual reckonings entail: "To rectify the noise in each sentence," she writes, "as I have done for the first three paragraphs, would render Lucy Church Amiably so rich in information as to
be incomprehensible, unless larger structures could be found" (1993, 186). However, all productive and rigorous interpretative activities are the result of this same sense of responsibility.

In "The World is Round," Stein alludes to the relationship between text, sound, and the performance space of reading that helps us understand how computational and musical adaptations are useful for reading her texts:

The teachers taught her
That the world was round
That the sun was round
And that they were all going around and around
And not a sound.
It was so sad it almost made her cry
But then she did not believe it
Because mountains were so high,
And so she thought she had better sing
And than a dreadful thing was happening
She remembered when she had been young
That one day she had sung,
And there was a looking-glass in front of her
And as she sang her mouth was round and was going
around and around.

Like the little girl who suddenly understands her position (her scary responsibility) in the universe, as the agent of not only the sound making but the meaning making, our interpretive activities as composers and coders, as producers and tool-developers, as musicians and as readers, as audience members, are guided by our responsibility to an ontological consideration for what makes a text knowable or understandable or interpretable and an epistemological consideration for how we know these interpretations are sound (pun intended). The notion that music "transcends referential or lexical meaning" is what attracted modernists to it as an artistic approach for representation (Bucknell 2001, 1), but it is the complementary notion that we cannot help but to make sense of what we are reading (to discover systems and larger structures) that makes it our responsibility to consider the role that musical and, I have argued by extension, computational adaptations play in our interpretive activities.

Works Cited


Distant Listening or Playing Visualisations Pleasantly with the Eyes and Ears by Tanya Clement. Sonic Imaginations, by John Sterne. Sample Datasets. Sonic Dictionary: a growing collection of more than 800 sound recordings created by university students to enhance the vocabulary of sonic experience. Penn Sound Archive: an archive of audio recordings, old and new. FMA: A Dataset for Music Analysis The FMA provides Creative Commons-licensed audio from 106,574 tracks from 16,341 artists and 14,854 albums, arranged in a hierarchical taxonomy of 161 genres. It provides full-length and high-quality data. What does all eyes and ears expression mean? Definitions by the largest Idiom Dictionary. Fig. listening and watching eagerly and carefully. Be careful what you say. The children are all eyes and ears. Tell us quick. We are all eyes and ears! See also: all, and, ear, eye. Want to thank TFD for its existence? Tell a friend about us, add a link to this page, or visit the webmaster's page for free fun content. Link to this page Distant Listening or Playing Visualisations Pleasantly with the Eyes and Ears. Tanya Clement. Published: 26 July 2013. by Open Library of the Humanities.